

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188		
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) 18-01-2011		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 8-Sep-2006 - 7-Sep-2010	
4. TITLE AND SUBTITLE SCIPUFF Tangent-Linear/Adjoint Model for Release Source Location from Observational Data			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER W911NF-06-C-0161		
			5c. PROGRAM ELEMENT NUMBER 206023		
6. AUTHORS O. O. ("Luwi") Oluwole, S.E. Albo, Richard C. Miake-Lye			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Aerodyne Research, Inc. 45 Manning Road Billerica, MA 01821 -3976			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 48892-CH-CDP.1		
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT The threat of atmospheric contamination by hazardous materials remains a high national security concern. There is a strong need for the development of emerging technologies which can significantly advance risk assessment and response capabilities. In this project, Aerodyne Research Inc. (ARI) has developed and demonstrated an					
15. SUBJECT TERMS Release Source Location, SCIPUFF, Inverse modeling, Threat detection, Source term estimation, Atmospheric dispersion modeling					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Richard Miake-Lye
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			19b. TELEPHONE NUMBER 978-663-9500

Report Title

SCIPUFF Tangent-Linear/Adjoint Model for Release Source Location from Observational Data

ABSTRACT

The threat of atmospheric contamination by hazardous materials remains a high national security concern. There is a strong need for the development of emerging technologies which can significantly advance risk assessment and response capabilities.

In this project, Aerodyne Research Inc. (ARI) has developed and demonstrated an algorithm for source estimation, called AIMS (“Aerodyne Inverse Modeling System”). AIMS takes as input all available observational data and optionally any prior knowledge of the source parameters. The output is the set of source parameters that best describes the observations, including number of sources, emission rates, locations, start and end times. AIMS is also designed to include an a posteriori assessment of its solution quality, providing useful feedback on how much confidence to put in a particular solution and in what ways the solution quality might be improved. A novel feature in AIMS is the ability to integrate multiple observation types in order to maximize information content for source estimation. This capability has been demonstrated for datasets from stationary and mobile sensors.

AIMS provides a significantly improved capability to protect the warfighter and civilians through successful identification and quantification of unknown hazardous atmospheric releases.

List of papers submitted or published that acknowledge ARO support during this reporting period. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Number of Papers published in peer-reviewed journals: 0.00

(b) Papers published in non-peer-reviewed journals or in conference proceedings (N/A for none)

Number of Papers published in non peer-reviewed journals: 0.00

(c) Presentations

1. SCIPUFF Tangent-Linear Adjoint Model for Release Source Location from Observational Data. Luwi O. Oluwole and Richard C. Miake-Lye, Chem-Bio Information Systems conference (Data Assimilation and Tactical Applications), 10 January 2007; Austin, TX
2. Source Estimation using SCIPUFF Tangent-Linear Adjoint Model, Luwi O. Oluwole and Richard C. Miake-Lye, 11th Annual GMU Atmospheric Transport and Dispersion modeling conference, 10 July 2007; Fairfax, VA
3. SCIPUFF Tangent-Linear and Adjoint Models for Release Source Estimation from Observational Data, Simón E. Albo, Luwi O. Oluwole and Richard C. Miake-Lye, 12th Annual GMU Atmospheric Transport and Dispersion modeling conference, 8 July 2008; Fairfax, VA
4. SCIPUFF-TLA : A plume inversion tool for identifying and quantifying pollutant emission sources from multi-measurement data, Oluwayemisi O. (“Luwi”) Oluwole, Simón E. Albo, Richard C. Miake-Lye, 13th GMU Conference, July 15 2009; Fairfax, VA
5. New developments in SCIPUFF Tangent-Linear Adjoint, Simón E. Albo, Oluwayemisi O. (“Luwi”) Oluwole and Richard C. Miake-Lye, 13th Annual GMU Atmospheric Transport and Dispersion modeling conference, 14 July 2009; Fairfax, VA

Number of Presentations: 5.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Oluwole, O.O., S.E. Abo and R. C. Miake-Lye, 2008; “Source Estimation Using SCIPUFF Tangent-Linear or Adjoint,” CBD Physical Science and Technology conference proceedings.

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

1

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

0

(d) Manuscripts

Number of Manuscripts:

0.00

Patents Submitted

Patents Awarded

Awards

Graduate Students

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Post Doctorates

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Under Graduate students supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale): 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields: 0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME

PERCENT SUPPORTED

FTE Equivalent:

Total Number:

Sub Contractors (DD882)

1 a. Sage Management

1 b. 6731 Columbia Gateway Drive

Suite 150

Columbia

MD

21046

Sub Contractor Numbers (c):

Patent Clause Number (d-1):

Patent Date (d-2):

Work Description (e): Developers of SCIPUFF. Provided technical assistance for implementation of inverse modeling tech

Sub Contract Award Date (f-1): 10/1/2006 12:00:00AM

Sub Contract Est Completion Date(f-2): 9/7/2009 12:00:00AM

Inventions (DD882)

SCIPUFF Tangent-Linear/Adjoint Model for Release Source Location from Observational Data

Final Report

ARO Contract No. W911NF-06-C-0161

September 8 2006 – September 7 2010

Prepared by

O. O. (“Luwi”) Oluwole, S.E. Albo, Richard C. Miake-Lye¹
Aerodyne Research, Inc.
45 Manning Road, Billerica MA 01821-397

Submitted to

US Army Research Office
PO Box 12211
Research Triangle Park, NC 27709-2211

¹ Principal Investigator; E-mail: rick@aerodyne.com; Phone: (978) 663-9500 x251

THIS PAGE INTENTIONALLY BLANK

Table of Contents

Project Objective.....	3
Background.....	4
SCIPUFF Atmospheric Transport and Dispersion Model.....	5
Prior Development.....	6
Benefit to Warfighter.....	8
Maturity of the Technology Prior to this Project.....	8
Project Accomplishments: The Aerodyne Inverse Modeling System (AIMS)	9
Overview.....	9
The Cost Function.....	9
Automatic Differentiation.....	11
“Forward” or “Tangent-Linear” Sensitivity.....	12
“Reverse” or “Adjoint” Sensitivity.....	12
Source Estimation Algorithm Heuristics	14
Quantifying Source Estimate Uncertainty	15
1. Solution Quality Index.....	15
2. Data Information Content Index	16
Modifications to SCIPUFF: Modeling Continuous Releases	17
Source Location Model Testing and Validation	19
Application to Ideal (model-generated) Data	19
AIMS Application to Field Data.....	28
Summary.....	33
Recommendations for Future Research & Development	34
References.....	35

Project Objective

The threat of atmospheric contamination by hazardous materials remains a high national security concern. There is a strong need for the development of emerging technologies which can significantly advance risk assessment and response capabilities. Thus, the objective of this project was the development and validation of SCIPUFF tangent-linear or adjoint model as a counterproliferation and counterforce tool for characterizing the source of a nuclear, biological, chemical (NBC) release using observational data. A fundamental aspect of gradient-based inverse modeling methods are the sensitivities of the ‘mismatch’ between numerical simulation and measurement for observables of a hazardous release to model parameters and inputs. Thus, another objective of this research project was to use the SCIPUFF tangent-linear or adjoint model to further the current understanding of the relationships between field observations, the characteristics of the release source, and the transport and dispersion of the hazardous cloud.

In this project, Aerodyne Research Inc. (ARI) has developed and demonstrated an algorithm for source estimation, called *AIMS* (“Aerodyne Inverse Modeling System”). AIMS takes as input all available observational data and optionally any prior knowledge of the source parameters. The output is the set of source parameters that best describes

the observations, including number of sources, emission rates, locations, start and end times. AIMS is also designed to include an *a posteriori* assessment of its solution quality, providing useful feedback on how much confidence to put in a particular solution and in what ways the solution quality might be improved. A novel feature in AIMS is the ability to integrate multiple observation types in order to maximize information content for source estimation. This capability has been demonstrated for datasets from stationary and mobile sensors.

AIMS provides a significantly improved capability to protect the warfighter and civilians through successful identification and quantification of unknown hazardous atmospheric releases.

Background

The standard modeling approach for the tracking of atmospheric plumes falls in the category of forward modeling: given an initial state (such as the atmospheric state, an initial tracer concentration), and, possibly time-dependent boundary values (such as space-/time-dependent tracer emissions and, in case of an offline meteorological background the time-evolving meteorological state), a transport model is stepped forward in time to produce a field of tracer concentrations at subsequent times.

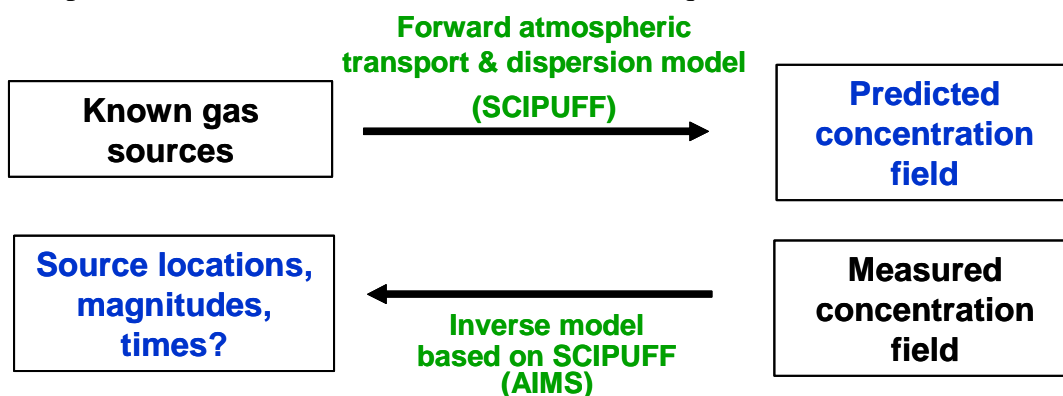


Figure 1. An illustration of atmospheric dispersion modeling in forward and inverse modes. In the forward mode, the gas source characteristics are well-known and the goal is to predict concentrations at a later time and downwind of the emission sources. In the inverse mode, downwind concentrations are known from measurements and the goal is to determine the source characteristics.

In source estimation, one wishes to compare the time-evolved tracer concentrations with a set of measurements of these quantities. This sort of problem falls in the category of inverse modeling – the goal is to determine the initial and/or boundary values that lead to a model trajectory most consistent with the observations. So, given the measurement data, one wishes to identify and quantify the emission sources. Among other benefits, inverse modeling enables more accurate assessment of the impacts of emissions events: unknown inputs (e.g. emission rates/quantities) required for the forward modeling can be automatically determined from measurement data using inverse modeling. Inverse atmospheric dispersion modeling is an area of active research [Bowen and Stratton

(1981), Daescu and Carmichael (2003), Errico (1997), Long et al. (2010), Lushi and Stockie (2010), Mavriplis (2007)].

The forward model of interest in this work is SCIPUFF (Second-Order Closure Integrated PUFF, [Sykes et al. (1985)]), which is a state-of-the-art atmospheric transport and dispersion model adopted by the Defense Threat Reduction Agency (DTRA). As described in the next section, SCIPUFF is equipped to model a variety of emission scenarios involving varying characteristics (locations, rates and times) and under wind conditions that vary spatially and temporally.

SCIPUFF Atmospheric Transport and Dispersion Model

The atmospheric dispersion model is the Second-order Closure Integrated Puff (SCIPUFF) model [Sykes et al., (1986, 1995, 1999)] which is the dispersion model employed in the Hazard Prediction and Assessment Code Capability (HPAC) [SAIC, 2001]. While SCIPUFF initially treated gaseous puffs evolving in the atmosphere, interactions with other materials are included in current versions. This includes solid particles, liquid droplets, and nuclear material and the ability to treat evaporation, deposition, and precipitation. Release sources include both continuous plumes and instantaneous puffs. A single calculation allows for any combination of continuous and instantaneous releases at any time throughout the temporal domain. Continuous releases are modeled as a series of single puffs releases one after another

The concentration field is modeled using the Gaussian puff method [Bass (1980)]. The concentration field is given by the summation of individual puffs, and the concentration for each puff, at a given time, is

$$c(\bar{x}) = \frac{Q}{(2\pi)^{\frac{3}{2}} \text{Det}^{\frac{1}{2}}(\sigma)} \exp \left[-\frac{1}{2} \sigma^{-1}_{ij} (x_i - \bar{x}_i)(x_j - \bar{x}_j) \right] \quad (1)$$

In Equation 1, Q is the puff mass, \bar{x} is the puff centroid, and σ is the puff spread. The dynamic relations are derived by moment methods, where the conservation equations are integration over all space. Turbulent diffusion is modeled by the second-order schemes of Donaldson [1973] and Lewellen [1977] where the second order fluctuation terms are solved for by a corresponding transport equation. The closure scheme provides a statistical variance of the concentration field, and the variance provides probability information. Thus, the SCIPUFF calculations provide mean concentration values together with a corresponding uncertainty estimate. Practical applications of the model on 50 km and 3000 km length scales have been performed [Sykes et al., (1986, 1995, 1999)].

Numerical techniques used include puff splitting, merging, and adaptive time steps. Because the puff centroid evolves based on the mean flow velocity at the centroid location, the Gaussian representation becomes less accurate as the puff grows. In order to overcome this problem, if a puff reaches a critical width, the puff is split into smaller puffs. The summation of small puffs can be used to accurately represent a generalized concentration field. As the smaller puffs evolve, if the puffs have a region of considerable overlap, then they are recombined or merged to a single Gaussian distribution. As different puffs evolve, each is exposed to a different dynamic environment, and as such,

different puffs may have different timescales for numerical integration. The adaptive time stepping scheme allows for individual puffs to be stepped at individual time step levels.

Gaseous materials released into the atmosphere in many cases have an initial vertical momentum and an initial temperature different than those of the surroundings. SCIPUFF has an option to include buoyancy effects in the calculations. The buoyancy effects are calculated with evolution equations instead of assigning the buoyant puff characteristics to the source. For instance, an effective release height is not used in the model. Instead the actual release height is used and subsequent change in puff centroid with altitude is calculated by solving the Boussinesq momentum equation and the conservation of potential temperature.

Meteorological input comes in a variety of forms: fixed winds (or uniform velocity vector), surface observational data, upper air observational data, and three-dimensional gridded data. Observational data includes mean wind, temperature, and possibly boundary layer parameters. Data is provided for a discrete number of locations and times. The information is interpolated over space and time to give the relevant quantity at the centroid of the puff of interest. An option for mass-consistency exists to enforce mass conservation with respect to the interpolated wind field.

Additional meteorological options include boundary layer and terrain. Boundary layer input parameters include surface roughness length, boundary layer depth, Monin-Obukhov length, and surface friction velocity. Given these parameters, the flow field in the planetary boundary layer is calculated. The relations are based on the modeling work of Wyngaard [1985], Venkatram and Wyngaard [1988], and Lewellen [1981]. The terrain option allows for a surface with varying height and properties. Surface elevation is read from a data file, and surface properties are set either based on the type of land cover (water, forest, cultivated, grassland, urban, desert) or are read from a data file.

Prior Development

Under a DTRA sponsored Phase II SBIR research project, automatic differentiation tools were used to develop the adjoint and tangent linear models for the SCIPUFF atmospheric transport and dispersion modeling component of Defense Threat Reduction Agency's Hazard Prediction and Assessment Capability (HPAC). Currently, the SCIPUFF adjoint model includes as controls the release source latitude, longitude, and height, as well as the release mass and release time. The model was tested using Dipole Pride 26 field data. For that data set, the observables were the measured tracer concentration downwind of the release from a set of whole air sampling bags located up to 20 km from the source.

Under the Phase II study, both (a) parameter optimization and (b) source location applications were explored. The distinction between the two applications is based on how much information is known about the specific control variables. For parameter optimization, the focus is on iterative adjustments in model input parameters which are known to within some degree of uncertainty in order to minimize the difference between model results and measured observables. For example, measured wind speed and

direction at selected locations would be examples of model input parameters which known to some extent, but which are subject to sufficient uncertainty to significantly impact model results. Parameter optimization could adjust these parameters within the uncertainty bounds to improve model simulations.

Figure 2 illustrates source location optimization using the SCIPUFF adjoint model for DP26 field tests as a scatter plot. The x-axis is the ratio of the predicted (optimized) source longitude to the report (actual) longitude. The y-axis is the ratio of the predicted (optimized) source latitude to the report (actual) latitude. Contours are also shown which delineate regions for which the distance between the predicted and reported source locations are less than some value in units of km. Thus, for most of the DP26 field trials, the predicted source location was within one kilometer of the actual release. Since the actual release locations are known, the data in Figure 2 means that differences between forward model calculations and measured release tracer concentrations result in predicted source locations which are within 1-2 km of the actual release.

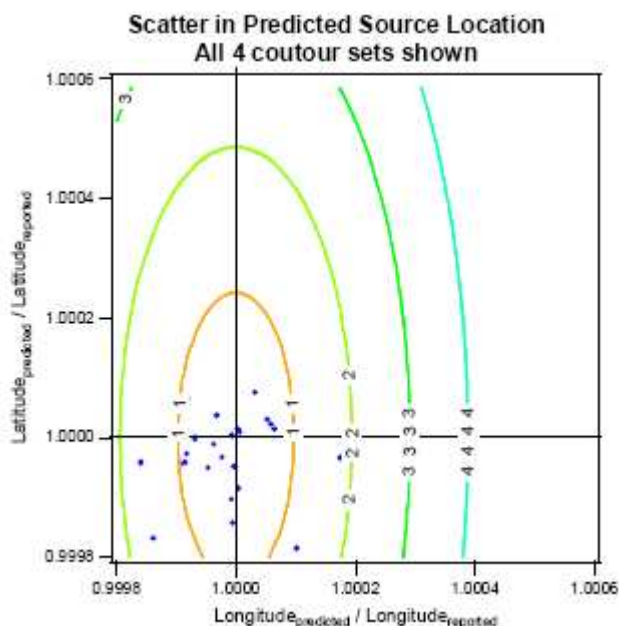


Figure 2: Scatter in SCIPUFF adjoint optimized source latitude and longitude for DP26 field data.

For the results in Figure 2, initial guesses for the source location were selected near the known release site. The SCIPUFF adjoint model was then used to iteratively adjust the source location to minimize the cost function. Locating the source of a hazardous release using field observations when there is no information about parameter values is more difficult. There are several potential approaches to this problem. One approach is a simple grid search for local minima for the cost function to map potential source locations. A second approach is to combine source optimizations as described above with results from other methods such as forward trajectory calculations which are being pursued by other groups. A third approach is an iterative optimization in which a set of initial parameter values are first selected and then optimized. The set of optimized parameters are then perturbed to define initial parameters for a second optimization. The

process is repeated leading to a map of local minima for the cost function through parameter space. This approach is illustrated schematically in Figure 3 in which the starting location is at a selected observational site or sensor location. This method was examined as part of the Phase II SBIR work and illustrative results for two of DP26 trials with results summarized in Table 1

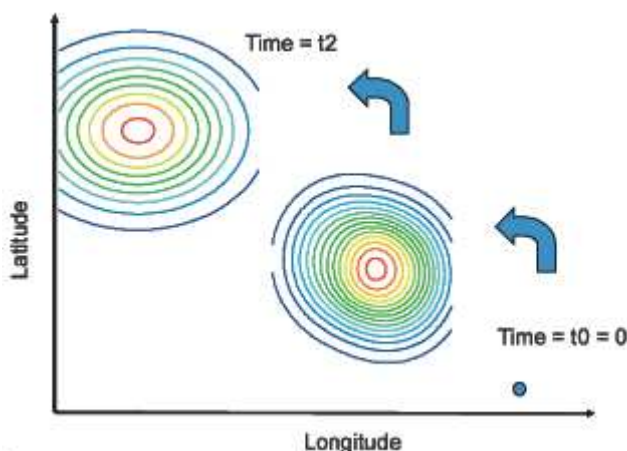


Figure 3. Schematic of iterative backtracking approach to source location.

Table 1. Summary of Predicted Release Source Parameters for DP26 Trials 11B.

Control Variable	Reported Value	Predicted Value	Absolute Difference
D26 Trial 11B			
Release latitude (E)	-116.0970	-116.0964	0.0006
Release longitude (N)	37.15986	37.1574	0.0025
Release time (UTC)	23.45	23.46	0.01

Benefit to Warfighter

The SCIPUFF inverse model would benefit the warfighter in two ways. The first is characterization of an unknown hazardous release source. The second is improved atmospheric transport and dispersion modeling capabilities through model parameter optimization.

Maturity of the Technology Prior to this Project

Prior to this project, a tangent-linear/adjoint for a simplified version of SCIPUFF atmospheric transport and dispersion model had been developed and tested using field data. However, for source location applications, the model was in a very early stage of development and a number of scientific and technical challenges needed to be addressed to enable further development for the desired (source location/quantification) applications.

Project Accomplishments: The Aerodyne Inverse Modeling System (AIMS)

Overview

The source of an atmospheric release can be described by using the following parameters: source location, source height, mass released, time of release and duration of release. Estimating a source refers to estimating all or some of these parameters. We have developed an advanced source estimation algorithm named *AIMS* (Aerodyne Inverse Modeling System). As illustrated in Figure 4, AIMS is based on SCIPUFF and takes as input all available measurement data and optionally any prior estimate of the source parameters. The output is the best set of source parameters for reproducing the measurement data using the forward model. Model outputs include number of sources, emission rates, locations, start and end times. AIMS is also designed to output an assessment of its solution quality, providing useful feedback on how much confidence to put in a particular solution and in what ways the solution quality might be improved (e.g. acquiring more measurement data).

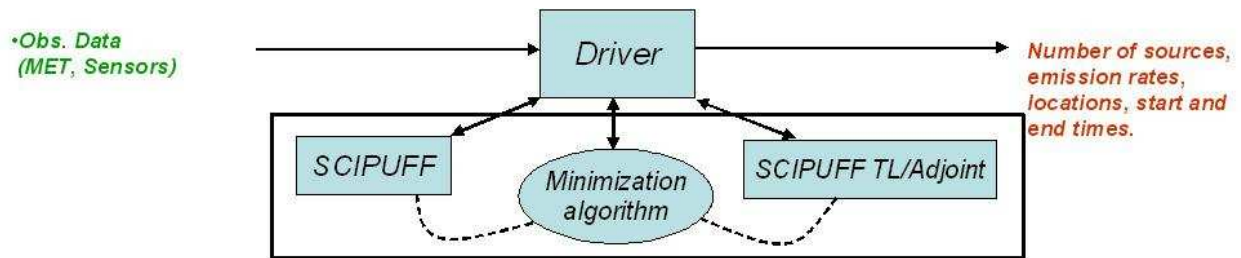


Figure 4: Schematic of the Aerodyne Inverse Modeling System (AIMS)

A novel feature in AIMS is its ability to integrate multiple observation types in order to maximize information content for source estimation.

The Cost Function

The quality of a source estimate can be evaluated by using a cost function to compare observed concentration data and model predictions when the source estimate is used in a forward calculation. Typically, the cost function is defined to be zero in the limit that model predictions match the data perfectly (i.e. when the model perfectly predicts the measurements at all sensors at all times), and increase as the quality of the source estimation worsens. The cost function developed and used in AIMS is as follows:

$$Cost = \sum_{m=1}^{N_{measurements}} \left\{ a_m \sum_{i=1}^{N_{times}} \frac{\sum_{j=1}^{N_{sensors}} [Y_{i,j}^{obs} - Y_{i,j}^{mod}(\beta)]^2}{\max \left\{ \sum_{j=1}^{N_{sensors}} (Y_{i,j}^{obs})^2, 0.05 * \max_{\forall i} \left(\sum_{j=1}^{N_{sensors}} (Y_{i,j}^{obs})^2 \right)_i \right\}} \right\} \quad (2)$$

where Y^{obs} refers to the observed concentrations and Y^{mod} refers to the model predicted concentrations at time i , at sensor j , obtained by running SCIPUFF (forward model) at a set of source parameters β ; and a_m is a weighting constant for each type of measurement. Note that the total cost in equation (2) is really a weighted sum of component cost functions m , for the different measurement types applied in the source estimation problem. This is to allow case-dependent specification of the desired relative impacts of the different observation types in a dataset. For instance, one may wish to include in a source estimation analysis stationary sensor measurements known to be highly uncertain, because the stationary sensors were at locations not covered by the more accurate, less noisy mobile concentration measurements. However, one wishes to do so without “over-contaminating” the source estimation with the large errors in the stationary measurement. This can be accomplished in AIMS by specifying appropriate weighting factors (e.g. $\alpha_m = 0.9$ and 0.1 for the mobile and stationary concentration measurements, respectively). Unless otherwise specified by the user, all measurement types are equally weighted in

AIMS; i.e. $\alpha_m = \frac{1}{N_{measurements}} \forall m$.

The current version of AIMS is able to treat both stationary and mobile concentration measurements. Our goal is to expand the capability to treat more observation types in future versions of AIMS.

Additionally, notice in Equation (2) that a minimum value for the denominator is set to be equal to 5% of the largest denominator. This is applied as a scaling mechanism to account for the potentially vast range of data values from the sensors over the course of the measurement period. In other words, applying this scaling avoids artificial inflation of the cost function by low-magnitude measurements at time i : notice that

as $\sum_{j=1}^{N_{sensors}} (Y_{i,j}^{obs})^2 \rightarrow 0$, the total cost is dominated by the model-data discrepancies at time

i , even for relatively low values of the numerator. On the other hand, omitting the denominator in Equation (2) causes the cost function to be dominated by the absolute model-data discrepancies in the larger-magnitude measurements, while the lower-magnitude data are essentially ignored. In that case, one would often obtain source estimates with large model-data discrepancies at the lower-magnitude data points. The scaling mechanism described above was found to be highly effective for simultaneously maximizing the information content of the wide range of data values typically found in practical observational datasets, for source estimation.

AIMS minimizes the cost function described in equation (2) using the Broyden-Fletcher-Goldfarb-Shanno algorithm [Press et al., (1992)], which uses a quasi-Newton method. The minimization algorithm requires multiple evaluations of the cost function, as well as its derivatives with respect to the unknown source parameters β . AIMS accomplishes this non-trivial task of obtaining the gradients of a complex software output with respect to its inputs by employing an Automatic Differentiation (AD) tool called TAF (Transformation of Algorithms in FORTRAN R. [Giering and Kaminski, (2006)]).

Automatic Differentiation

Automatic Differentiation allows evaluation of the derivatives of a function specified by a computer program. Note that this is not a trivial matter, since one often does not have a closed form expression relating outputs to inputs in numerical models; and determining these relations is typically intractable, particularly in large, complex models like SCIPUFF. AD tools take as input the source codes that constitute the forward model and automatically generate corresponding source codes for computing the gradients.

We apply the AD tool called TAF [Giering and Kaminski, (2006)]. TAF exploits the chain rule for computing the first derivative of a function with respect to a set of input variables. Treating a given forward code as a composition of operations – each line representing a compositional element, the chain rule is rigorously applied to the code, line by line. The resulting tangent linear or adjoint code returns the Jacobian matrix computed in forward (following model evaluation steps from input to output) or reverse (backtracking from output to input) order, respectively. Both approaches yield accurate analytical gradients, though the tangent-linear is more straight-forward to implement; it is also more computationally efficient when computing gradients of several outputs to only a few inputs. The adjoint mode is more efficient for computing gradients of a few outputs to several inputs [Zhang et al., (2008)]. In practice, the adjoint mode involves some additional computational overhead costs and only outperforms the tangent-linear when the number of outputs is much greater than the number of inputs.

A little linear algebra helps to understand the issue. Let M be a general, nonlinear model, i.e. a mapping from the m -dimensional space $U \subset \mathbb{R}^m$ of input variables $\bar{u} = (u_1, \dots, u_m)$ (initial conditions such as concentrations, boundary conditions such as emissions, or model parameters) to the n -dimensional space $V \subset \mathbb{R}^n$ of model output variable $\bar{v} = (v_1, \dots, v_n)$ (model concentrations or diagnostics thereof, objective functions such as a model vs. data misfit, ...) under consideration,

$$\begin{aligned} \underline{M} : \underline{U} &\rightarrow \underline{V} \\ \underline{u} &\rightarrow \underline{v} = \underline{M}(\underline{u}) \end{aligned} \tag{3}$$

The vectors $\bar{u} \in U$ and $\bar{v} \in V$ may be represented with respect to some given basis vectors $\text{span}(U) = \{\bar{e}_i\}_{i=1, \dots, m}$ and $\text{span}(V) = \{\bar{f}_i\}_{i=1, \dots, n}$ as

$$\vec{u} = \sum_{i=1}^m u_i \vec{e}_i, \quad \vec{v} = \sum_{j=1}^n v_j \vec{f}_j$$

Two routes may be followed to determine the sensitivity of the output variable \vec{v} to its input \vec{u} : forward (or direct sensitivity) and reverse (or adjoint sensitivity).

“Forward” or “Tangent-Linear” Sensitivity

Consider a perturbation to the input variables $\delta \vec{u}$ (typically a single component $\delta \vec{u} = \delta u_i \vec{e}_i$). Their effect on the output may be obtained via the linear approximation of the model M in terms of its Jacobian matrix M , evaluated in the point $u^{(0)}$ according to

$$\delta \vec{v} = M|_{u^{(0)}} \delta \vec{u} \quad (4)$$

with resulting output perturbation $\delta \vec{v}$. In components $M_{ji} = \partial M_j / \partial u_i$, it reads

$$\delta v_j = \sum_i M_{ji}^{(0)} \delta u_i \quad (5)$$

Eq. (4) is the tangent-linear model. In contrast to the full nonlinear model M , the operator M is just a matrix which can readily be used to find the forward sensitivity of \vec{v} to perturbations in u , but if there are very many input variables, it quickly becomes prohibitive to proceed directly as in (4), if the impact of each component e_i is to be assessed. As an example, let v_j be the concentration at a point x_j at the end of the model integration, \vec{u} a field of initial emissions, and ask what is the sensitivity of v_j to u_r , i.e. what is the change δv_j under changes $\delta \vec{u}$. To obtain a full picture, one would have to perturb each component of the initial field δu_i , $i=1, \dots, m$ separately and perform a forward calculation for each perturbed state.

“Reverse” or “Adjoint” Sensitivity

Let us consider the special case of a scalar objective function $J(\vec{v})$ of the model output. This scalar is either the concentration at a certain model grid point as in the above example, or a measure of some model-to-data misfit. The description of the model as in Eq. (3) can then be extended to read

$$J: U \rightarrow V \rightarrow R$$

$$\vec{u} \rightarrow \vec{v} = M(\vec{u}) \mapsto J(\vec{u}) = J(M(\vec{u})) \quad (6)$$

The perturbation of J around a fixed point J_0 , $J = J_0 + \delta J$ can be expressed in both bases of \vec{u} and \vec{v} with respect to their corresponding inner product

$$\begin{aligned}
J &= J|_{\vec{u}} + \langle \nabla_{\vec{u}} J^T |_{\vec{u}}, \vec{\delta u} \rangle + O(\|\vec{\delta u}\|^2) \\
&= J|_{\vec{v}} + \langle \nabla_{\vec{v}} J^T |_{\vec{v}}, \vec{\delta v} \rangle + O(\|\vec{\delta v}\|^2)
\end{aligned} \tag{7}$$

After some straightforward algebra the gradient $\nabla_{\vec{u}} J$ can be readily inferred by invoking the adjoint M^* of the tangent linear model M

$$\begin{aligned}
\nabla_{\vec{u}} J^T |_{\vec{u}} &= M^T |_{\vec{u}} \cdot \nabla_{\vec{v}} J^T |_{\vec{v}} \\
&= M^T |_{\vec{u}} \cdot \vec{\delta v} \\
&= \vec{\delta u}
\end{aligned} \tag{8}$$

Eq. (6) is the adjoint model, in which M^T is the adjoint (here, the transpose) of the tangent linear operator M , $\vec{\delta v}^*$ the adjoint variable of the model state \vec{v} , and $\vec{\delta u}^*$ the adjoint variable of the control variable \vec{u} .

The reverse nature of the adjoint calculation can be readily seen as follows. Consider a model integration which consists of A consecutive operations $M_A(M_{A-1}(\dots(M_1(M_0(\vec{u}))))))$, where the M 's could be the elementary steps, i.e. single lines in the code of the model, or successive time steps of the model integration, starting at step 0 and moving up to step A , with intermediate $M_\lambda(\vec{u}) = \vec{v}^{(\lambda+1)}$ and final $M_A(\vec{u}) = \vec{v}^{(A+1)} = \vec{v}$. Let J be a cost function that explicitly depends on the final state \vec{v} only (this restriction is for clarity reasons only). $J(u)$ may be decomposed according to:

$$J(M(\vec{u})) = J(M_A(M_{A-1}(\dots(M_1(M_0(\vec{u})))))) \tag{9}$$

Then, according to the chain rule, the forward calculation reads, in terms of the Jacobian matrices (we've omitted the $|$'s which, nevertheless are important to the aspect of tangent linearity; note also that by definition $\langle \nabla_{\vec{v}} J^T, \vec{\delta v} \rangle = \nabla_{\vec{v}} J \cdot \vec{\delta v}$)

$$\begin{aligned}
\nabla_{\vec{v}} J(M(\vec{u})) &= \nabla_{\vec{v}} J \cdot M_A \cdot \dots \cdot M_1 \cdot M_0 \cdot \vec{\delta u} \\
&= \nabla_{\vec{v}} J \cdot \vec{\delta v}
\end{aligned} \tag{10}$$

whereas in reverse mode we have

$$\begin{aligned}
 M^T(\nabla_v J^T) &= M_0^T \cdot M_1^T \cdot \dots \cdot M_A^T \cdot \nabla_v J^T \\
 &= \nabla_u J^T
 \end{aligned}
 \tag{11}$$

clearly expressing the reverse nature of the calculation.

Source Estimation Algorithm Heuristics

Figure 5 summarizes the source estimation algorithm implemented in AIMS. Several heuristics were developed to improve the algorithm robustness, including:

1. Automatic starting guess heuristic – automatically determines an appropriate starting guess of the unknown source parameters, so that users are not required to have any prior knowledge/guess of source parameters. This algorithm works as follows: it starts by identifying the maximum concentration measured; its corresponding measurement time and location are used as the initial guess, together with a low mass value (10^{-10} Kg).
2. Cost function minimization heuristics – to separate the cost minimization with respect to release time from its minimization with respect to the other unknown source parameters (source location, mass and duration); we found this approach to improve the source estimation success rate. This algorithm works as follows: after evaluating the cost for the initial guess, two other earlier release times (determined automatically) are also evaluated. Thereafter, the time search will be expanded on the direction of the lowest cost until a minimum is bounded. Once a minimum is bounded, the time intervals are bisected until the desired time interval precision is achieved. The best estimated source is the one that yields the lowest cost function value. When dealing with multiple sources, the above procedure is repeated up to three additional times (total of four sources), each time adding a starting guess for a new source to the solution from the previous round. To determine the number of sources, the optimal cost for the different guesses are compared and the final solution is the total number of individual sources that yielded the lowest-magnitude cost.

Additionally, the cost function is artificially enlarged by addition of a constant whenever a non-physical solution is encountered in the minimization process, to force the algorithm back within the range of physically viable source parameter values. This enabled us to essentially create a constrained minimization algorithm while retaining the (unconstrained) quasi-Newton solver which we have found to be very effective for AIMS. Parameter values classified in AIMS as non-physical are: (a) negative source mass; (b) source estimates at locations undetectable by the sensors (in the case of multiple sources); and (c) negative source duration (for continuous sources).

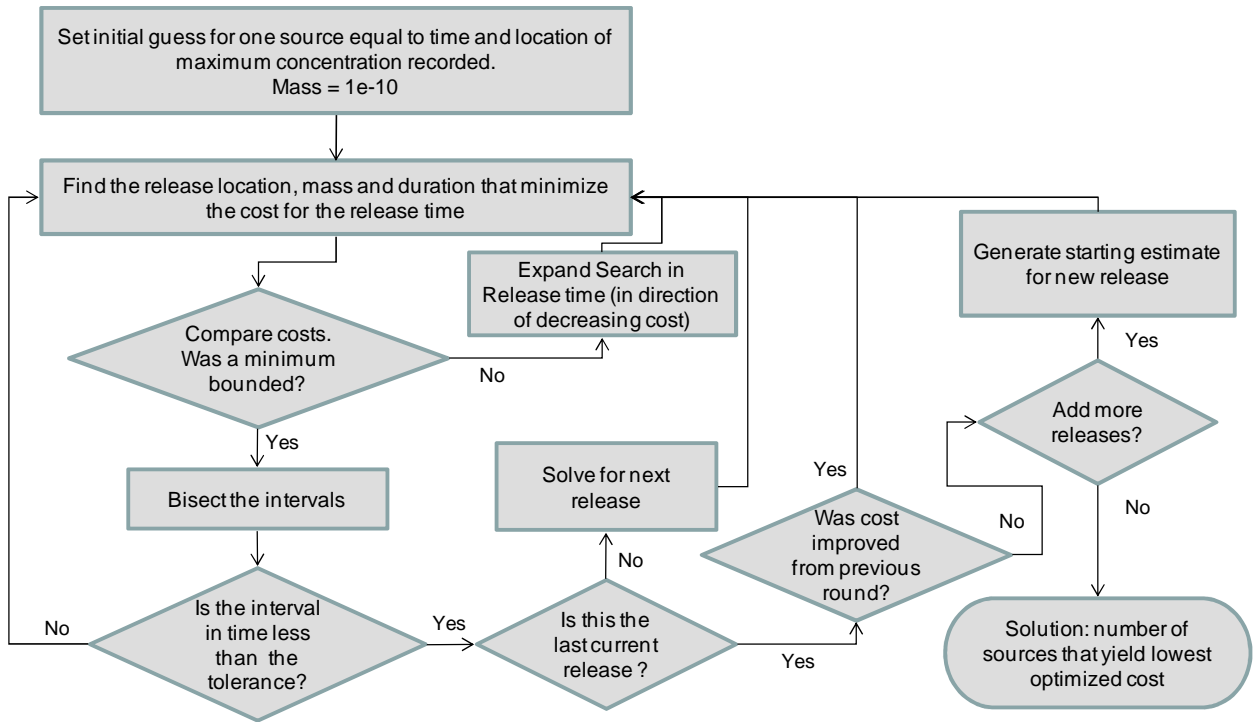


Figure 5: Summary of source estimation algorithm implemented in AIMS

Quantifying Source Estimate Uncertainty

Uncertainties in the final source estimate can arise from two main sources: 1) imperfect plume inversion; and 2) non-ideal data and dispersion (forward and inverse) modeling. These uncertainties are quantified using the following metrics:

1. Solution Quality Index

Given ideal data and perfect dispersion model, the accuracy of the source estimate is limited only by the success of the plume inversion; i.e. the magnitude of $Cost(\beta^*)$ in Equation (2). For the exact solution, $Cost(\beta^*) = 0$; therefore, the magnitude of this term quantifies the error in the final solution. Although it is not trivial to exactly quantify the error in β^* from the error in $Cost$, it is clear that they are directly proportional. So, a larger cost value indicates a larger inversion error.

In practical source estimation scenarios, it may be difficult to determine the accuracy of the source estimate, since the actual source parameters are not known. However, it is clear that solutions with larger cost values are generally less accurate than those with smaller costs. Therefore, we define a simple *solution quality index*, as described below, for quantifying the accuracy of source estimations using AIMS.

It is interesting to note the behavior of the cost function when the sources are moved away from the observation regions. For instance, consider a modeled release occurring at a location or time such that it is not detected by the sensors. In this case

$Y_{i,j}^{\text{mod}}(\beta) = 0 \quad \forall i, j$; and $\text{Cost} = \text{const} \equiv \text{Cost}_{\infty}$ (see equation (2)). It is clear that this is an inaccurate source estimate and, as described earlier, AIMS considers this an inadmissible solution. Thus, it is convenient to define the solution *quality* index as:

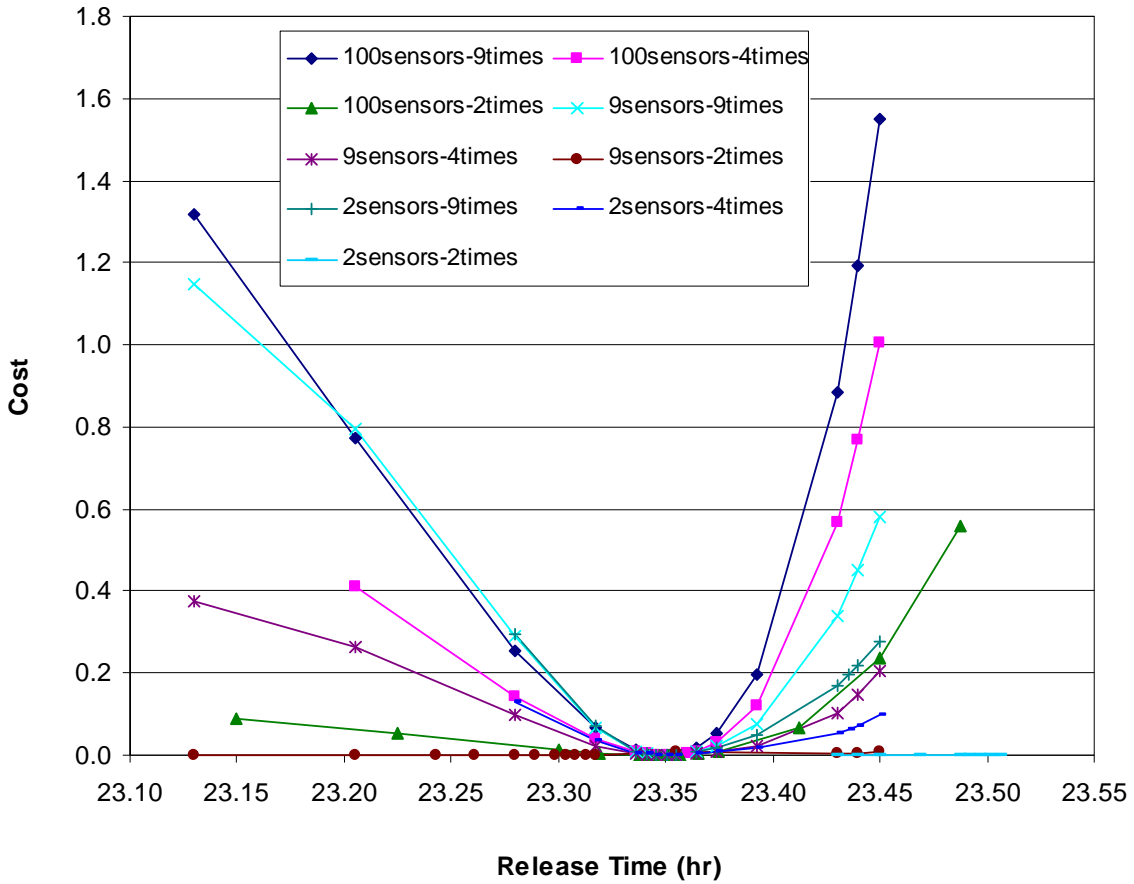
$$\phi = 1 - \frac{\text{Cost}(\beta)}{\text{Cost}_{\infty}}, \quad (12)$$

such that $\phi = 1$ for a perfect source estimation involving ideal data and model, $\phi = 0$ for inadmissible solutions as described above, and in general $0 \leq \phi \leq 1$. In addition to imperfect source estimation, other factors that contribute to non-unit ϕ are data errors and model errors (i.e. forward dispersion model's inability to accurately represent reality). A useful interpretation for ϕ is that it is proportional to the fraction of the total information in the measurement data that is accurately captured by the forward model Y^{mod} evaluated at the source estimate β .

2. Data Information Content Index

As described earlier, the uniqueness of an inverse problem solution relies on “sufficient” information in the observations. Without sufficient data information content, it is possible to have several solutions with perfect feasibility, $\phi = 1$. However, one cannot rely on such estimates, as they are highly uncertain.

Again, it is useful to consider the behavior of the cost function in relation to data information content. For simplicity, we consider a single point concentration sensor recording *useful* observations over a given time period. As used here, *useful* observations are those that capture the temporal evolution of the passing plume at a given location; in other words, the rise in concentration, followed by the plateau and the decay. Addition of other such sensors at different locations provides increasing information content on the spatial evolution of the plume and decreases the space of highly feasible solutions for Equation (2), since more constraints must now be satisfied. So, the cost function well becomes narrower with increasing data information content. This is illustrated with synthetic data on the following figure. A similar concept has been observed by Sharan et al. [2009].



These findings motivated the definition of a *data information content* index:

$$\frac{\phi(\beta^*) - \lambda}{t_{rel}^* - t_{rel}|_{\phi=\lambda}} * \frac{L_c}{\bar{u}}, \quad (13)$$

where λ is a user-set change in feasibility (we use 0.1), L_c is a characteristic length scale (we use the distance between a source and the nearest sensor), \bar{u} is the average horizontal wind velocity and t_{rel} is the estimated release start time. One interpretation for this metric is that it quantifies the change in ϕ over a characteristic time interval, defined here as the approximate travel time from source to nearest sensor. A larger value denotes better information content.

Modifications to SCIPUFF: Modeling Continuous Releases

SCIPUFF forward model is capable of simulating a continuous release from a source, however there are several difficulties for operating TAF on this section of the code. The difficulties are rooted in the data management and embedded computational syntax used

and create discontinuities for the application of the chain rule by the automatic differentiation code.

In order to successfully include continuous releases in the inverse model one possible approach would have been to identify and construct workarounds for all incompatibilities, this could have potentially been very time intensive. Another possible solution is to start from TAF-compatible code for instantaneous releases and reproduce SCIPUFF continuous release algorithm avoiding known TAF-unfriendly constructs. The second approach was chosen and it is based in using a series of consecutive instantaneous puff releases that spread and overlap downstream approximating a continuous source, as shown in Figure 6. The algorithm implementation details, which highlight the differences between SCIPUFF original code and the TAF-friendly implementation, are presented in Figure 7.

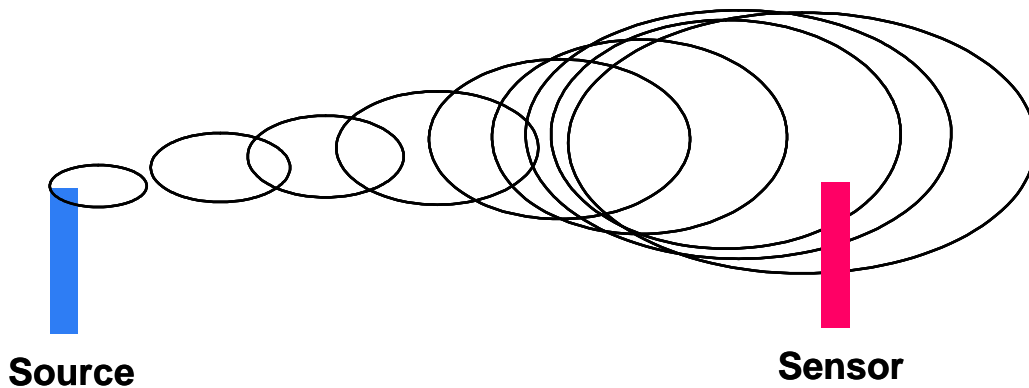


Figure 6: Continuous source approximated using a series of consecutive instantaneous puff releases

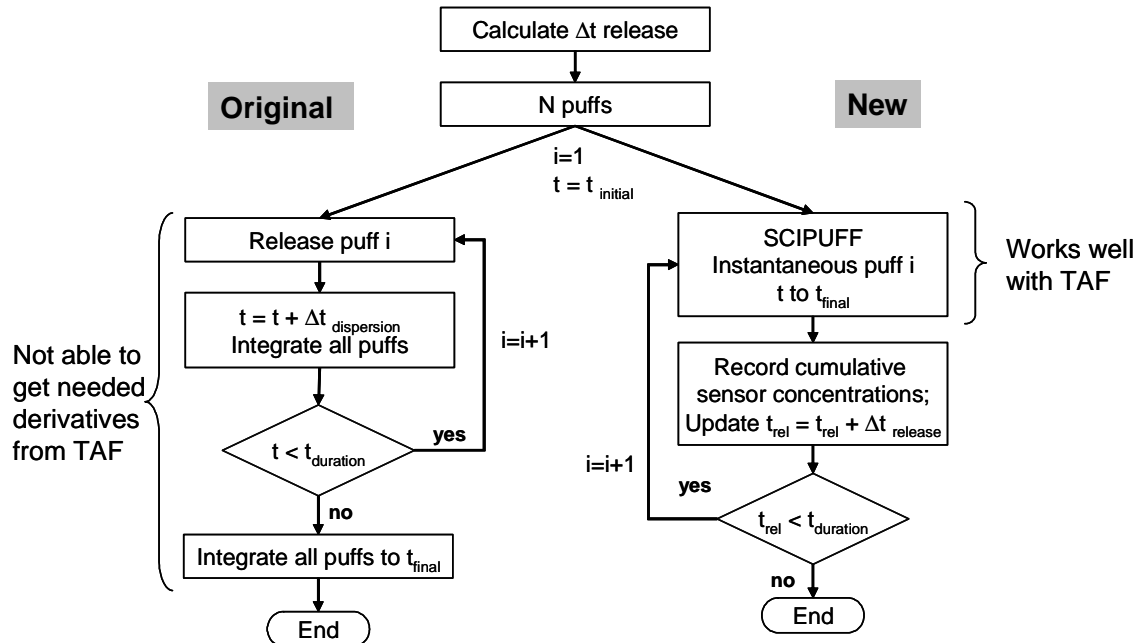


Figure 7: Algorithm implementation for modeling continuous releases

The TAF-friendly implementation replaces the complex metrics used in the original code for determining time interval between puff releases with a heuristic rule. A comparison between the predictions of both codes shows that the new AIMS implementation reproduces the predictions of the original code.

The methods described above have been implemented in AIMS and have been successfully applied for several source estimation scenarios, with illustrative results presented in the next section.

Source Location Model Testing and Validation

AIMS has been successfully applied to several problems for identification and quantification of emission sources. A series of applications to model-generated data enabled performance testing which was used for algorithm development and validation. Subsequently, AIMS has also been successfully used in a number of field data applications.

Application to Ideal (model-generated) Data

In this section, we report the performance of AIMS for several test cases involving single-source and multiple-source events, as well as stationary and mobile concentration measurement data. In each case, we consider both instantaneous and continuous release types. Finally, we investigate the effects of data quantity and quality by varying the number and spatial resolution of sensors and by artificially adding random noise to model-generated sensor data.

The release-detection configurations used in the following test cases were based on the initial plans for the FUSION Field Trials performed in September of 2007 (FFT-07) [Platt et al. (2008)]. The basic setup consists of 100 concentration sensors evenly distributed over a 1-km square grid (see Figure 8).

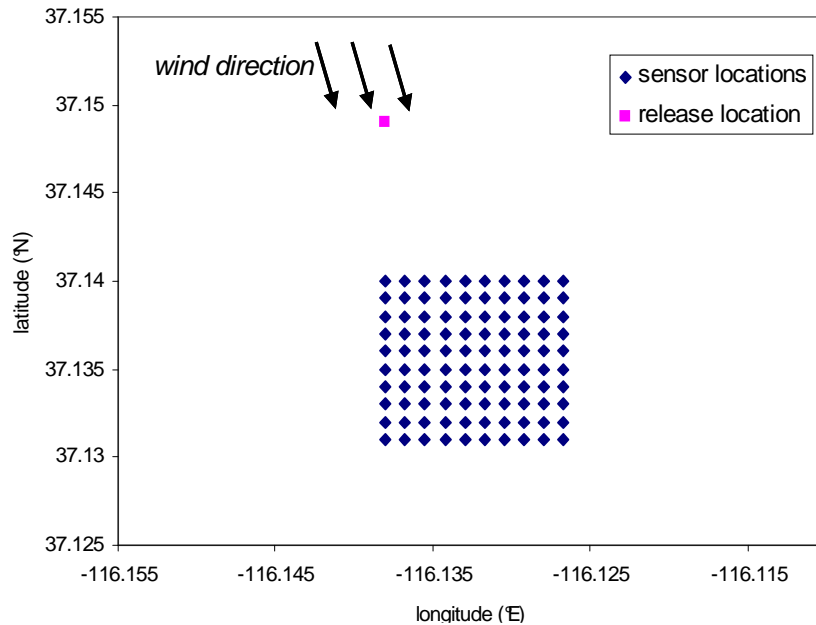


Figure 8: FFT-07-inspired configuration showing stationary sensor network, wind direction and release location.

We simulated observational sensor data by recording SCIPUFF-computed concentrations at specified measurement times and positions, assuming one or more releases of SF₆ (chemically inert) gas. We also assumed uniform and constant meteorological conditions over the release-detection domain. In each case, we applied AIMS to estimate the (assumed unknown) corresponding source parameters (location(s), time(s), mass(es)/rate(s), duration(s), number of sources). Note that by using model-generated data, we are able to control secondary effects like model and data uncertainties in evaluating the performance of our algorithm.

1. Single-source scenarios

As a simple starting point, we tested the method for cases where one knows that only one source is involved.

Instantaneous releases

The release parameters to be estimated are position, time and mass. We applied a constant and uniform two-dimensional wind field with a velocity of 3.2m/s and direction 346 degrees. At these conditions, we simulated a 100g release, 3m above ground at position (-116.38°, 37.149°) and we observed the concentration profiles shown in Figure 9. We then recorded observational data for each sensor at eight 3-minute time intervals, starting at 6 minutes after the release occurred.

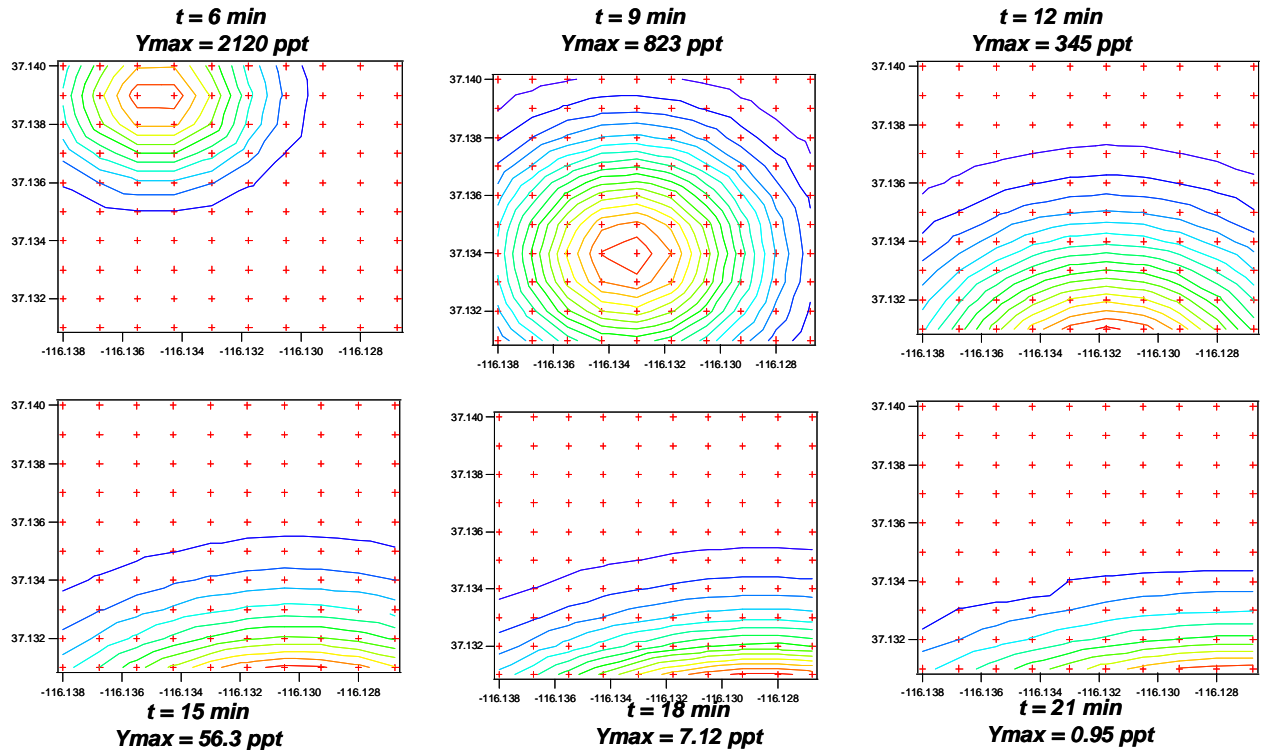


Figure 9: SCIPUFF-generated concentration profiles illustrating data sampling process. The grid markers indicate stationary sensor positions, where sample data were recorded. Only the first six sampling times are shown because they capture the puff entry into and exit from the sensor network.

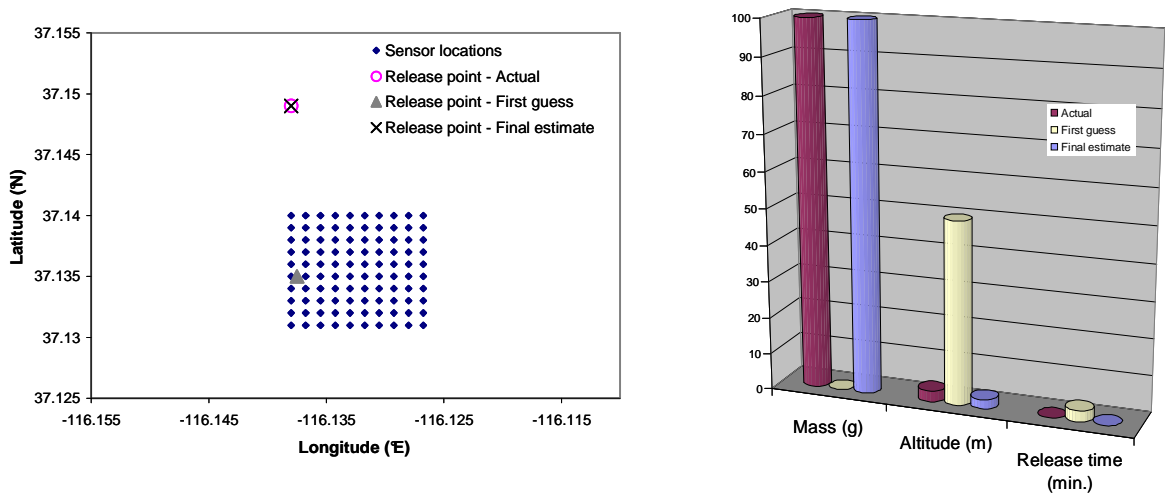


Figure 10: Source estimation results for a single instantaneous release.

Figure 10 shows that our source estimation algorithm very accurately estimates the unknown source in this problem, even with very poor starting guesses. In fact, the source is identified exactly in this case. Note that we report release altitudes relative to ground level and release times relative to the actual release time.

Continuous releases

Increasing the complexity of the test problem, we applied AIMS to a continuous-source case. In addition to the release time, position and strength estimated for the unknown instantaneous release, here we must also estimate the unknown release duration.

We applied the same meteorological conditions and generated observational data at the same sampling times described for the previous test case. Again, assuming an unknown source, we applied AIMS to estimate the source parameters, given the observational data. The results in Figure 11 demonstrate that we are able to accurately estimate unknown sources involving a single continuous release.

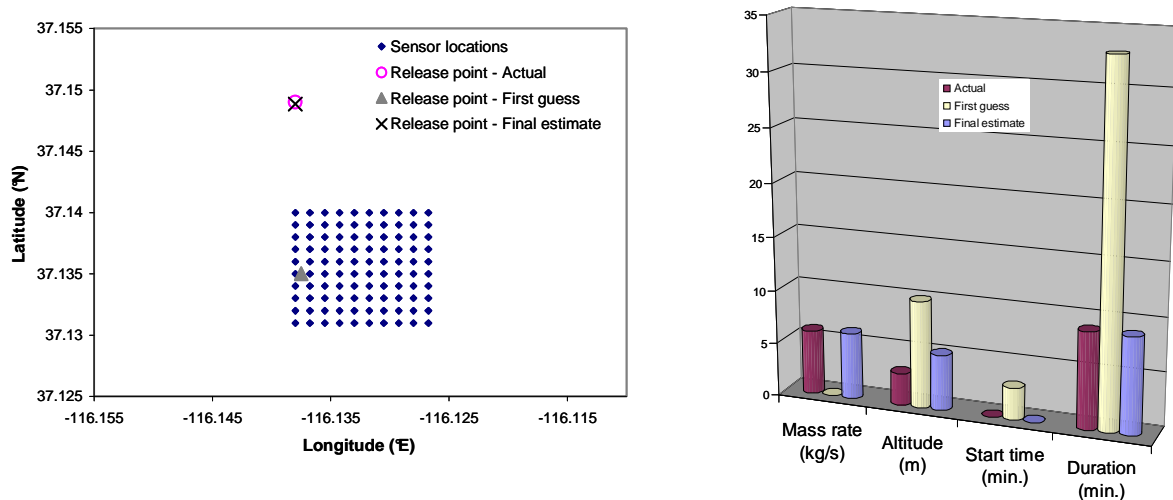


Figure 11: Source estimation results for a single continuous release.

2. Multiple-source scenarios

In practice, source estimation problems can involve multiple sources contributing simultaneously to the observational dataset. In this subsection, we investigate the performance of our algorithm for such scenarios. As usual, we will first consider the simpler case of instantaneous releases before we treat continuous releases. When dealing with multiple sources, one must estimate an additional parameter – the number of sources. For simplicity, we first assumed that all releases in these tests are known to have occurred at the same time. We found that, unlike in the single-source cases, here we needed to assume known release altitudes in order to obtain satisfactory source estimates. More recent advances in the algorithm development have eliminated the need for this restriction, as will be seen later in the sections on field data application.

For the test problems discussed in the current section, note that although we assumed all other source parameters to be unknown, for clarity we have reported source location estimates only. Source location accuracies were representative of the performance for the other source parameters.

Instantaneous releases

We considered release scenarios involving 2 or 4 sources. Figures 12 to 15 show the source estimates assuming 1, 2, 3, or 4 releases in each case, and the cost function value

for each estimate. Figure 12 and Figure 14 represent cases where the observational data clearly indicated multiple separate events, while in Figure 13 and Figure 15 all puffs overlapped significantly over the sensor network and it was not clear from the concentration profiles that there were multiple releases. Measurement times and meteorological conditions were identical to those applied for previous cases, except in Figure 13 where we applied a 0-degree wind direction.

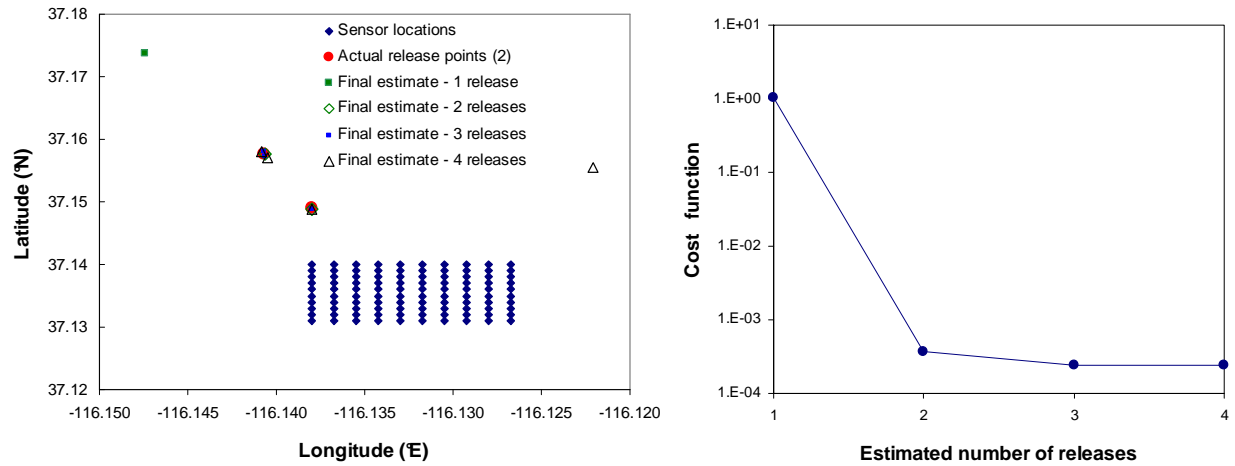


Figure 12: Source estimation results for two simultaneous instantaneous releases – minimal puff overlap.

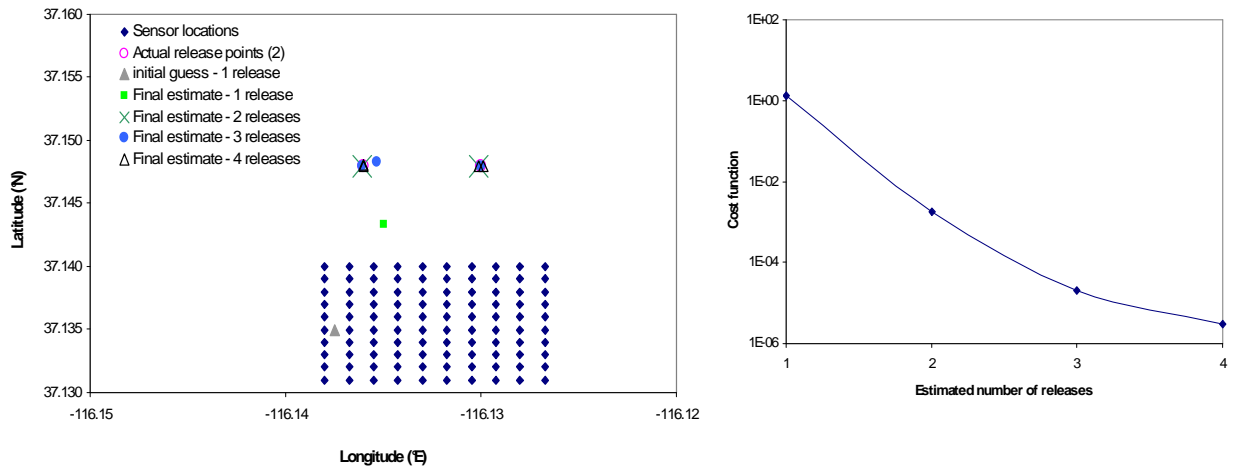


Figure 13: Source estimation results for two simultaneous instantaneous releases – significantly overlapping puffs.

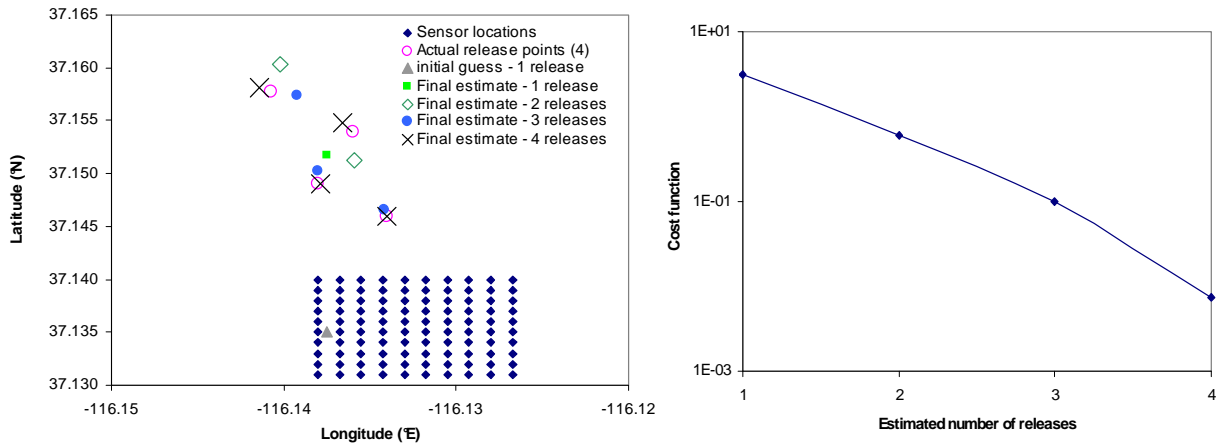


Figure 14: Source estimation results for four simultaneous instantaneous releases – minimal puff overlap.

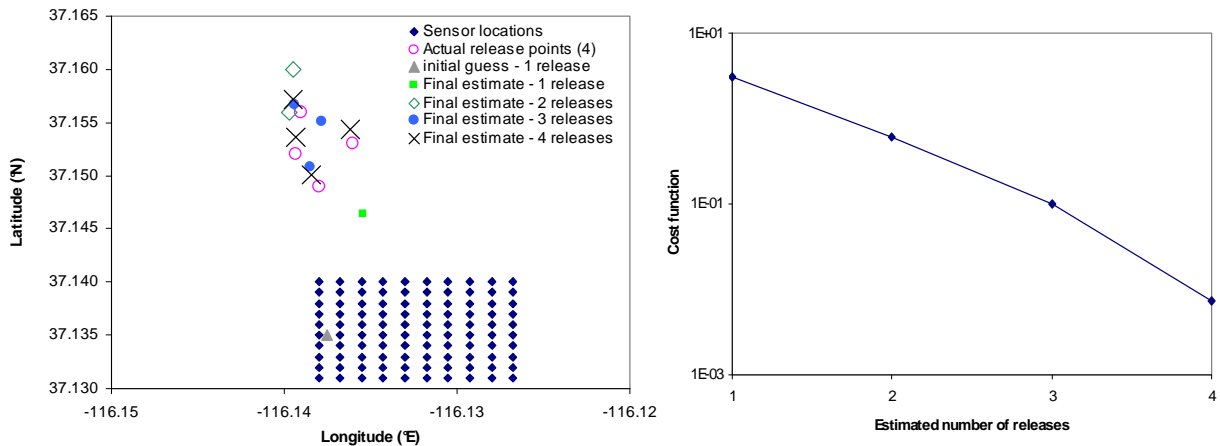


Figure 15: Source estimation results for four simultaneous instantaneous releases – significantly overlapping puffs.

In general, our algorithm correctly identified the numbers and locations of all sources. Not surprisingly, cases involving puff overlaps or more release points were more challenging. Note in Figure 13 that the cost function was lowest when we assumed four sources, even though only two sources were actually present. Closer examination reveals that the method estimated two identical releases at each of the two actual release points and accurately estimated the total masses – effectively accurately describing the two actual sources. Although the result is ultimately accurate, we are surprised by this behavior and we hope to gain a better understanding through some ongoing studies.

Continuous releases

We repeated the numerical experiment described in Figure 13, this time applying two continuous sources of equal duration. The result is shown in Figure 16. Here, we found that we also needed to assume known release durations (in addition to release altitudes) in

order to obtain satisfactory source estimates. Subsequent advances in the algorithm development eliminated this limitation, as will be seen later when field data application is discussed.

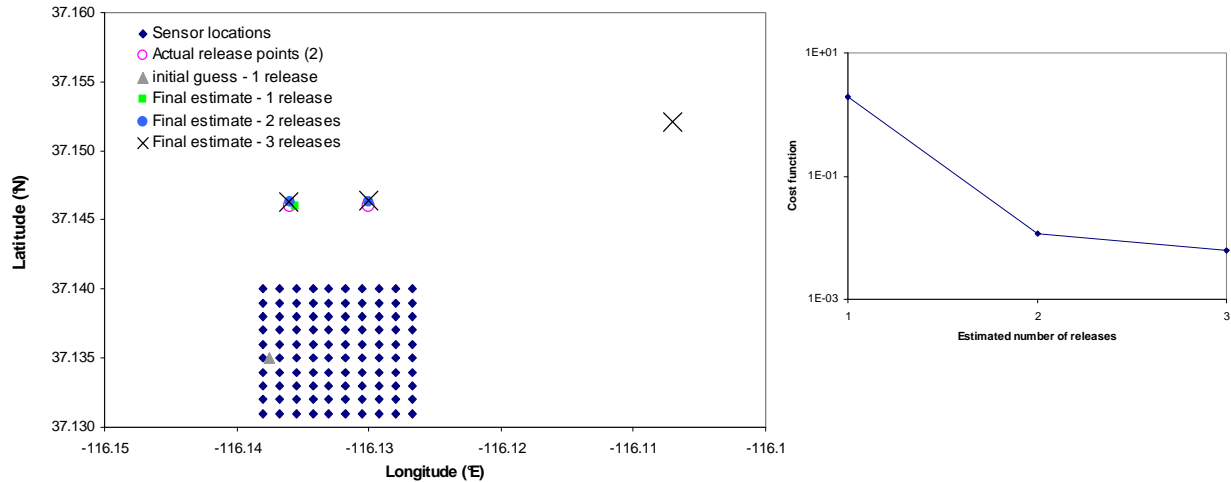


Figure 16: Source estimation results for two simultaneous continuous releases – significant plume overlap.

Another interesting observation here is that although there were actually only two releases, the cost function was lowest for the three-release configuration. However, one of the releases was positioned such that it essentially never impacted the sensors – effectively (and accurately) describing two sources.

3. Effects of data quantity

All of the cases considered so far have applied observational data from 100 sensors in a 1-km square grid. In practice, one would most often have much fewer sensors, with lower spatial resolution. To investigate the impact of such a decline in data quantity on the performance of AIMS, we repeated the single-source numerical experiments described earlier, gradually decreasing the number of sensors.

Figure 17 and Figure 18 show that AIMS is robust to low data quantities. Not surprisingly, continuous sources appear to require more sensor data for accurate source estimation than do instantaneous sources.

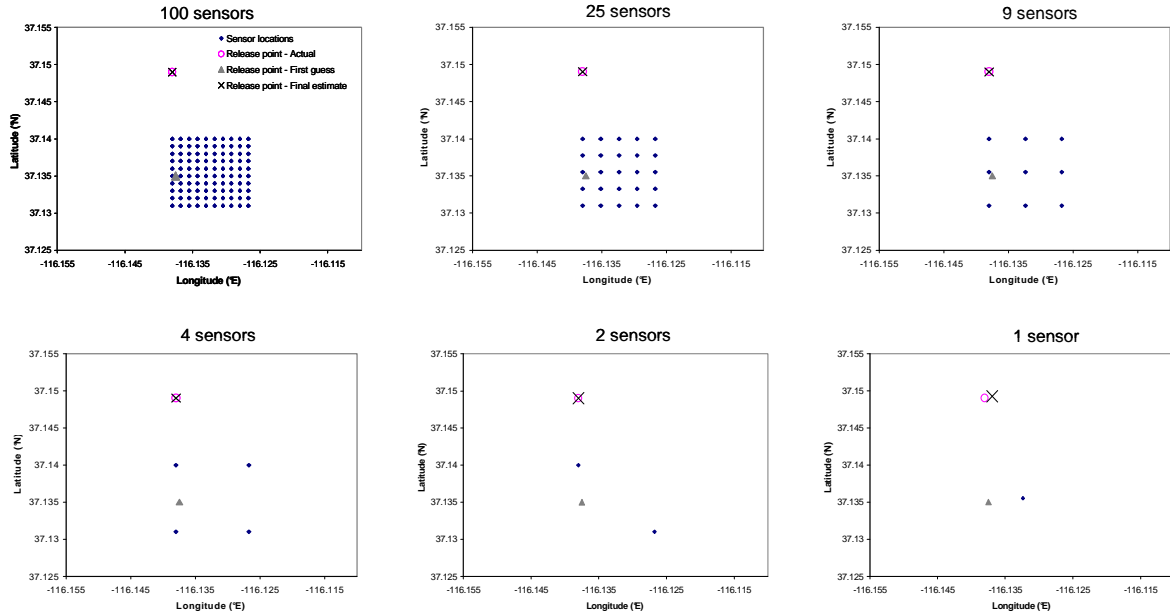


Figure 17: Source estimation results for single instantaneous source – gradually decreasing data quantity.

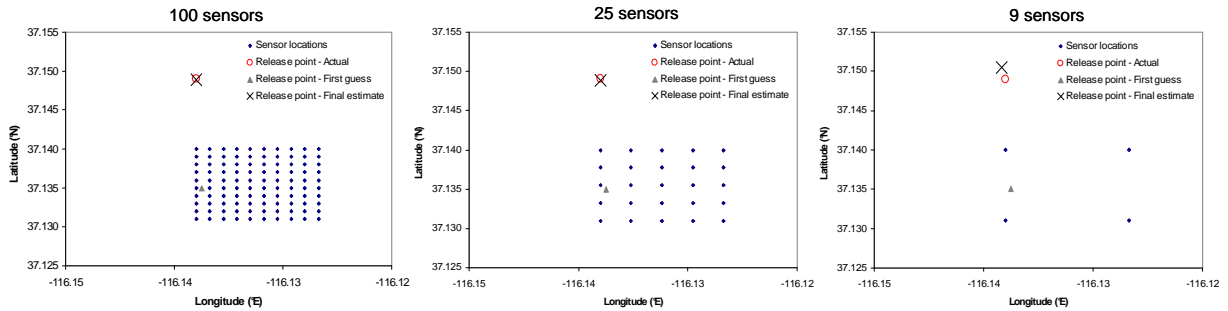


Figure 18: Source estimation results for single continuous source – gradually decreasing data quantity.

4. Effects of data quality (noisy sensor data)

Finally, we investigated the impact of data quality on the performance of AIMS. Since all sensor data were model-generated, we artificially added random noise to simulate measurement uncertainties. We applied random errors from a normal distribution in the range $\pm(3\%$ of actual concentration value + 15) ppt, which is representative of measurement uncertainties in practice. AIMS accurately estimated the sources using 100, 25, and 9 sensors, and the impact of the noise was first observed with 4 sensors (see Figures 19 and 20).

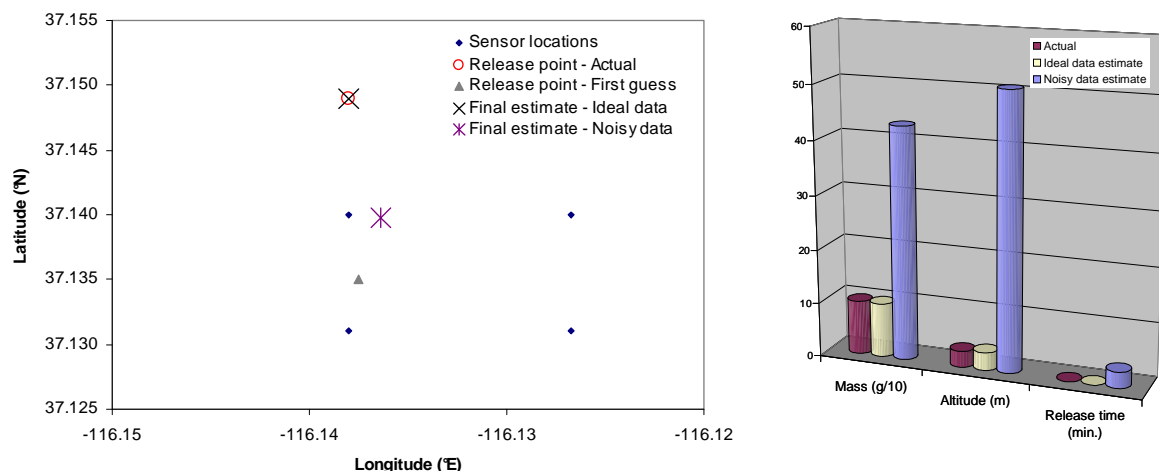


Figure 19: Source estimation for single instantaneous source – ideal versus noisy sensor data.

These results indicate that our source estimation method is robust to noisy data, but is more sensitive to data quantity with noise than without; perhaps more data are needed to “regularize” the effects of noise.

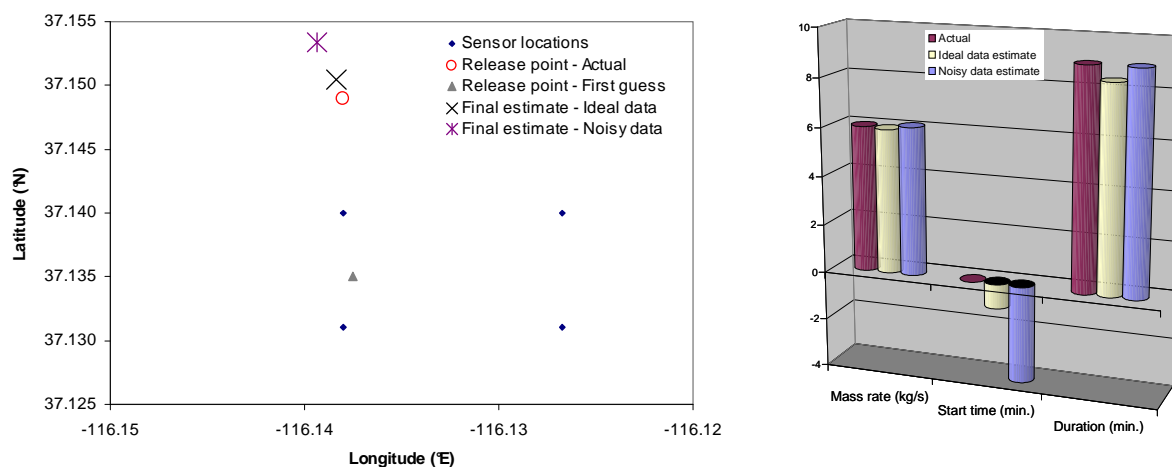


Figure 20: Source estimation for single continuous source – ideal versus noisy sensor data.

5. Multi-Measurement Data: Beyond Stationary Concentration Sensors

All of the previous examples involved concentration measurements using stationary sensors. However, one of the goals of AIMS is to enable integration of multiple measurement types in order to maximize information content for source estimation. In addition to stationary concentration measurements, the current version of AIMS is able to process mobile concentration measurements for source estimation. This is demonstrated in Figure 21.

In this example, mobile sensor data are obtained by simulating a mobile instrument recording ground-level propene concentration data while driving around a 4-km region surrounding several industrial facilities. The goal is to identify the source of propene in

this scenario, assuming that the actual source is unknown. As shown in Figure 21, AIMS successfully identifies the source, using data from the mobile concentration measurements.

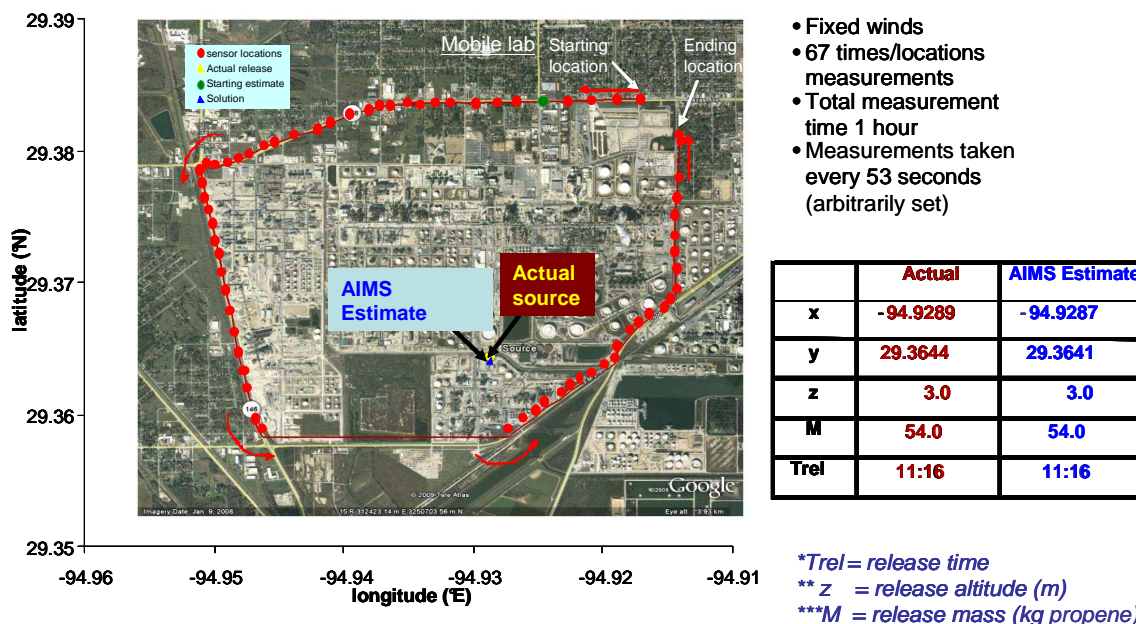


Figure 21: AIMS successfully identified the unknown source in this example involving mobile concentration measurements.

AIMS Application to Field Data

In addition to model-generated data shown in the previous examples, AIMS has also been successfully applied to field data, for controlled-release (experimental) scenarios as well as for real unknown source scenarios. The results of these AIMS applications are discussed below.

FFT-07 Controlled Releases

A field trial campaign, called *FUsing Sensor Information from Observing Networks (FUSION) Field Trial 2007* (or “FFT-07”) was performed in 2007, funded by the Defense Threat Reduction Agency (DTRA). FFT-07 was a short-range (approximately 1-2 km) controlled dispersion test designed to collect data to support development of prototype algorithms for source estimation. A range of release scenarios were considered: daytime and nighttime emissions, single and multiple sources, instantaneous and continuous sources. The details of FFT07 are provided in Platt and Deriggi [2010]. In summary, the data provided to algorithm developers involved 104 cases of propene gas releases, with 4 or 16 stationary sensors used for data sampling in each case. The unknowns were: number of sources, source locations, strengths, times and durations. AIMS was demonstrated as a robust and reliable model, consistently in the top rankings for all scenarios in an independent inter-comparison study by the Institute for Defense Analyses involving eight inverse plume dispersion algorithms (Platt et al. 2010). On average, AIMS source estimates were within 100m of the actual sources. These results are summarized on the figures below.

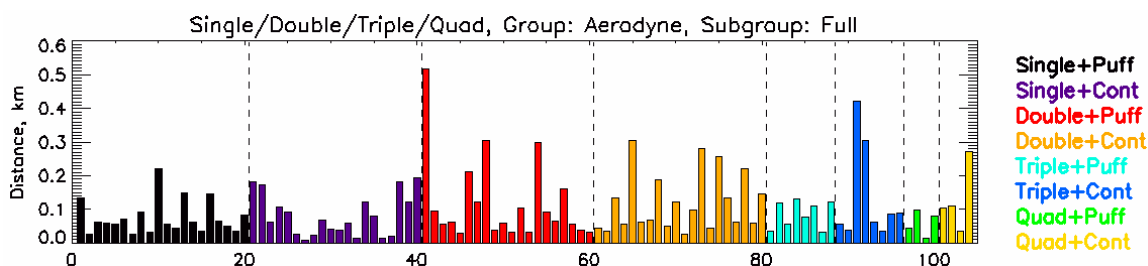


Figure 22: Distance between AIMS-estimated source location and actual FFT07 source location, averaged over all sources in each case. Results are grouped by number of sources and source types.

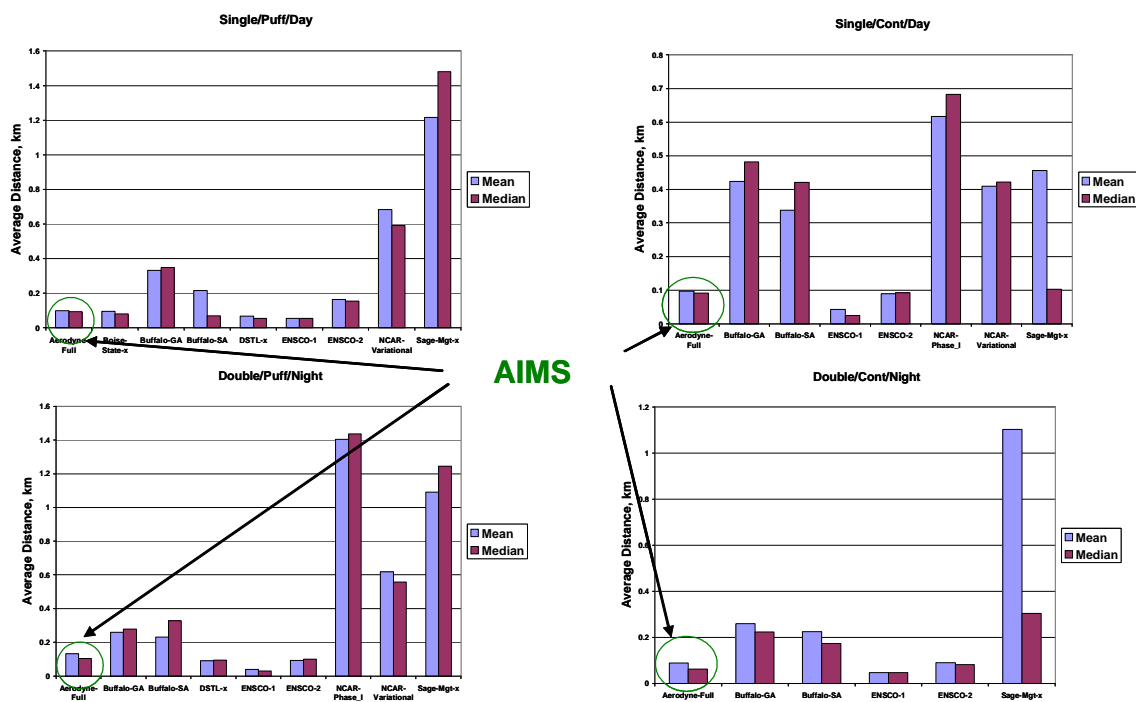


Figure 23: Average distance between estimated source location and actual FFT07 source location for different participating algorithms. The four plots correspond to different types of releases: Single day puff, single day continuous source, double night puffs, double night continuous sources. Smaller bars indicate better predictions as the estimated location is closer to the real location.

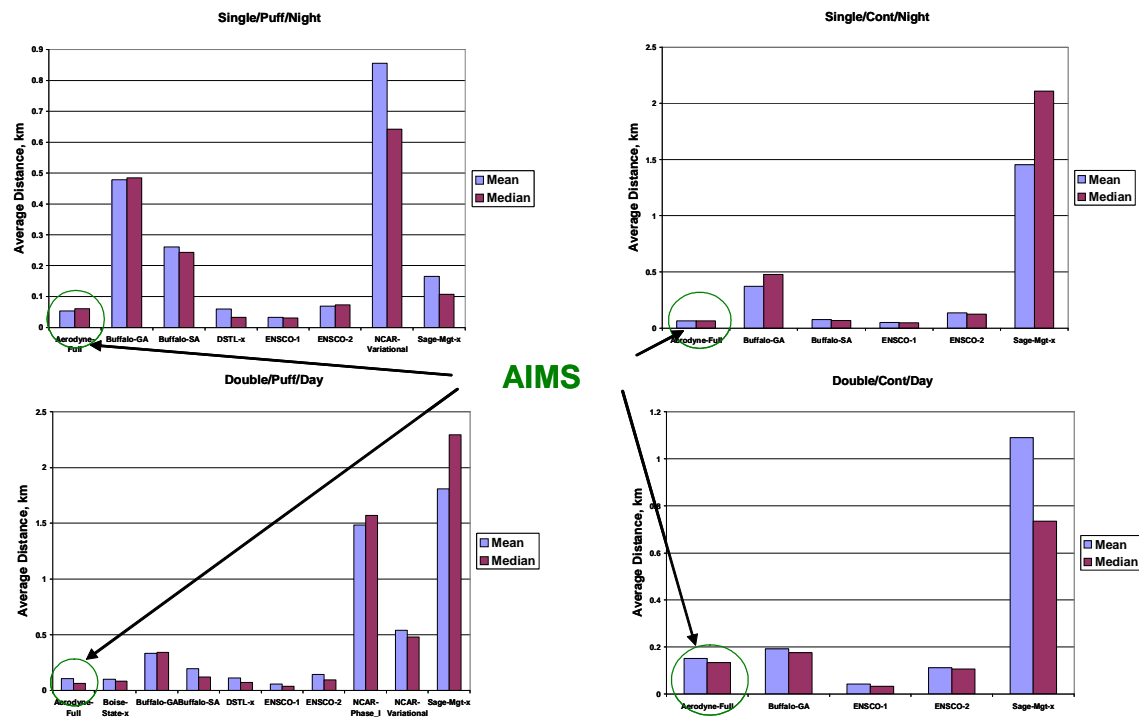


Figure 24: Average distance between estimated source location and actual FFT07 source location for different participating algorithms. The four plots correspond to different types of releases: Single night puff, single night continuous source, double day puffs, double day continuous sources. Smaller bars indicate better predictions as the estimated location is closer to the real location.

In this application, the best results were obtained for stable wind conditions and the most challenging runs involved highly varying winds or narrow sensor signals, as illustrated in the following figures.

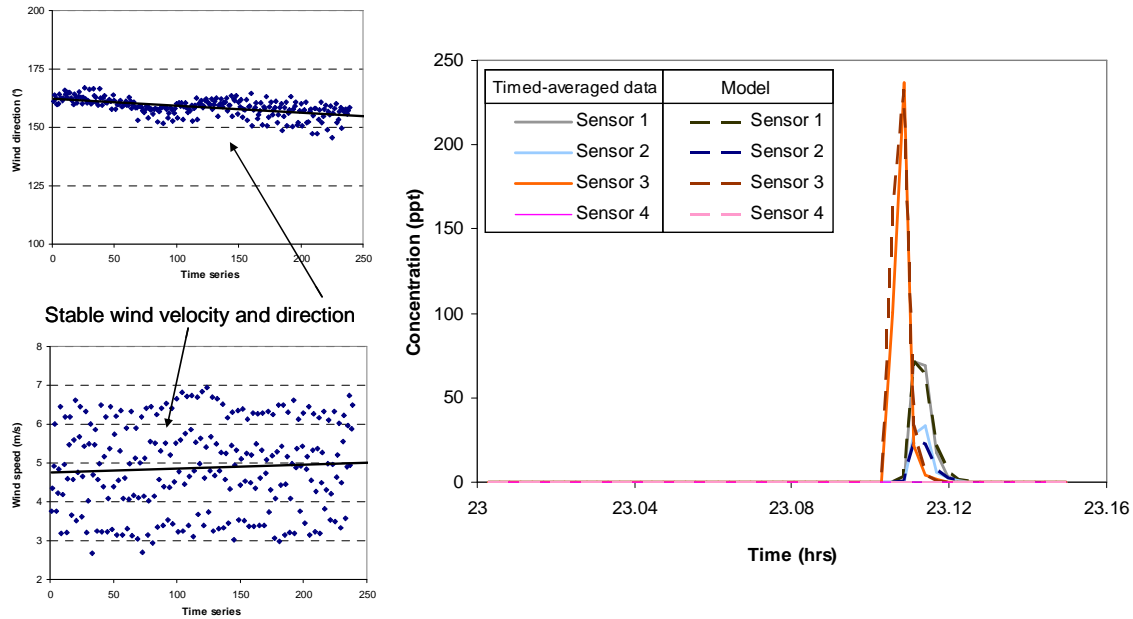


Figure 25: Example FFT07 case with stable wind velocity and direction (Case 26). Good agreement is shown between model and data concentration profiles. The estimated solution approximates well the real release parameters.

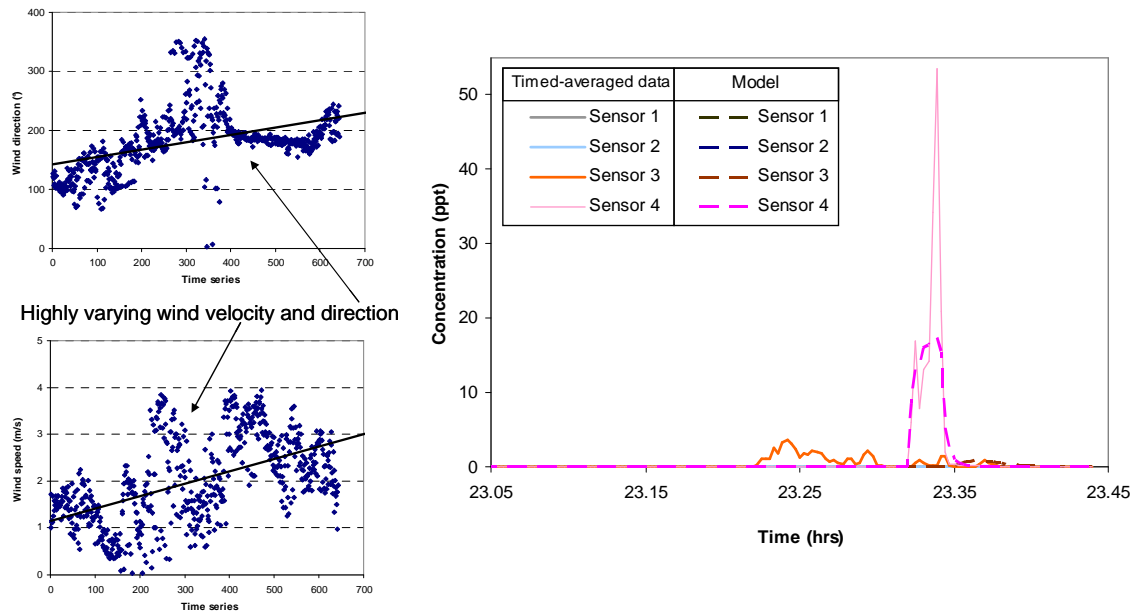


Figure 26: Example FFT07 case with highly variable wind velocity and direction (Case 47). Agreement between model and data concentration profiles is not optimal. The estimated solution might not approximate the real release parameters as well as in the case of stable winds.

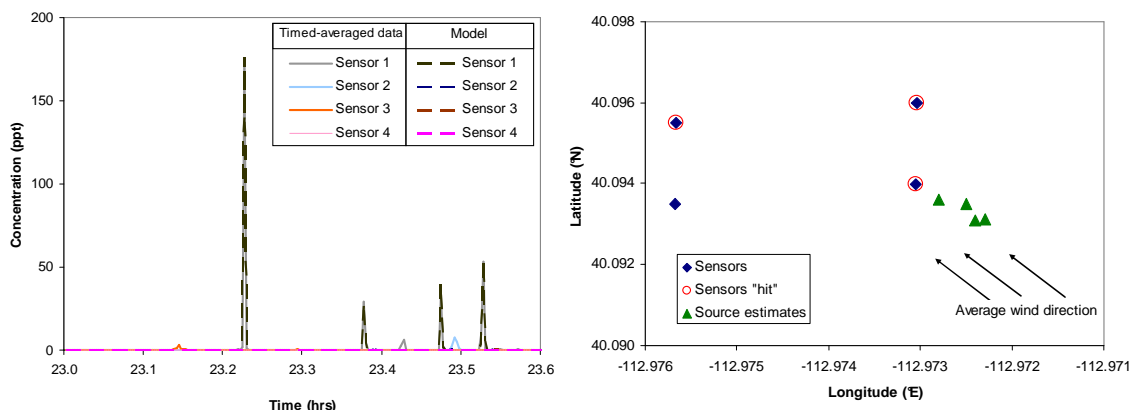


Figure 27: Example FFT07 challenging case with narrow concentration signals (Case 41). Narrow signals indicate different release times for each puff. A good initial guess for location and release time of each event are very important because the cost penalty for not predicting a peak with the calculated solution is not too high (only a few concentration points in the peak).

The Study of Houston Atmospheric Radical Precursors (SHARP)

SHARP was a large multi-organizational air quality analysis campaign performed in 2009, with one of its goals being identification and quantification of key air pollutant sources. Along with other instruments, the Aerodyne Mobile Laboratory (pictured below) was used for continuous sampling, to monitor the air quality in order to detect unwanted industrial process emissions.



Figure 28: The Aerodyne Research Inc. Mobile Laboratory.

The ARI mobile laboratory is outfitted with many instruments, several of which are developed and built at ARI using proprietary technology, for obtaining high-resolution, real-time measurements of gases, particles and aerosols [Kolb et al. (2004)]. An onboard anemometer records wind speed and direction data. A significant advantage over stationary sensors is that one can obtain a wealth of useful and relevant data with a single sensor by following observed (non-lethal) plumes with the mobile lab (with a tailored truck path).

Note that this source estimation problem is complicated by the fact that there is a high density of industrial activity in the Houston area; therefore there are several potential sources. AIMS was successfully used to determine the emission rates, locations and times of measured combustion-source pollutants in the Houston area. The solution quality index defined earlier (equation 12) was used to evaluate the source estimation solutions in these cases, since the real sources were unknown. Solution qualities for all AIMS estimates were no less than 75%. Figure 29 shows an example application of AIMS with mobile concentration measurements during the SHARP campaign.

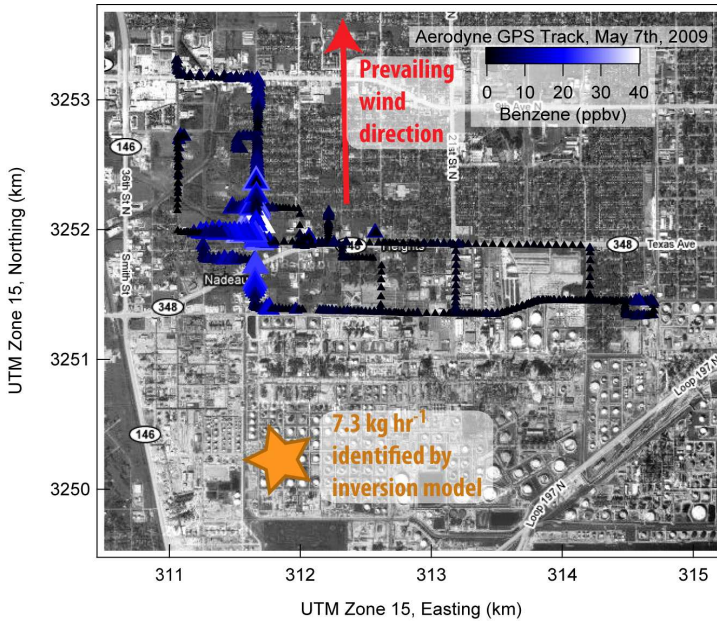


Figure 29: Aerodyne mobile laboratory measurements of benzene (blue triangles) and the responsible emission source magnitude and location (yellow star) identified using AIMS. The observed wind direction is indicated by the red arrow. The measurements were taken on May 7, 2009 in Texas City during the Houston SHARP campaign.

Summary

Aerodyne Research Inc. (ARI) has developed and demonstrated an algorithm for source estimation, called *AIMS* (“Aerodyne Inverse Modeling System”). In general terms, AIMS applies a variational approach for source estimation: a cost function is defined that quantifies the mismatch between all observations and the corresponding model predictions resulting from a given set of trial source parameters; then, the optimal set of source parameters is identified as the values for which the cost function is minimized:

$$\begin{aligned} \text{Cost}(\beta) &= \| \text{Data} - \text{Model}(\beta) \| \\ \beta^* &= \arg \min_{\beta} \text{Cost}(\beta) \end{aligned} \quad (14)$$

where β is the set of unknown source parameters; and β^* is the value of β that yields forward model predictions that are most consistent with the data.

Indeed, in the theoretical limit of ideal data and models, the global minimum of this cost function exists at the set of parameters that is most likely responsible for the observational data. The two main challenges of variational approaches in practice involve successfully locating the (global) minimum of the cost function and dealing with non-ideal data and models. The former challenge demands careful definition of the cost function and the use of a robust minimization algorithm. The latter requires awareness of (and accounting for) artificial offsets in the location of the minimum, due to non-ideal data and models.

AIMS takes as input all available observational data and optionally any prior knowledge of the source parameters. The output is the set of source parameters that best describes the observations, including number of sources, emission rates, locations, start and end times. AIMS is also designed to include an *a posteriori* assessment of its solution quality, providing useful feedback on how much confidence to put in a particular solution and in what ways the solution quality might be improved. A novel feature in AIMS is the ability to integrate multiple observation types in order to maximize information content for source estimation. This capability has been demonstrated for datasets from stationary and mobile sensors.

AIMS provides a significantly improved capability to protect the warfighter and civilians through successful identification and quantification of unknown hazardous atmospheric releases.

Recommendations for Future Research & Development

1. Sensor technologies have evolved and continue to evolve beyond point concentration data. Expanding the capability of source location algorithms to accommodate new types of observations has the potential to not only increase their applicability but also their accuracy since more available information can be integrated for source estimation. This capability can be implemented in AIMS in the future.
2. Recent advances in computational science and technology have created new opportunities for code parallelization that can be exploited to reduce source location algorithm run times and enable more rapid, real-time threat assessment. In particular, graphics processing units (GPUs) formerly reserved for computer graphics rendering have evolved into highly cost-efficient tools for highly parallel scientific computation. Several components of the source location algorithm developed here involve large amounts of computations that can each be performed independently and would be well-suited for parallelization. This feature can be implemented in AIMS in the future to increase its speed of source location.
3. Many real-life release scenarios involve time-dependent sources, either in mass flow rate, position or both. Expanding the capability of source estimation algorithms like AIMS to characterize time-varying sources would be very valuable.

4. AIMS can be extended to automatically provide guidance on optimal sensor placement. This would be especially relevant for those cases when the data information content index indicates that the source estimates are highly uncertain. This new capability would be used to plan sensor locations in preparation for potential events.

References

C. T. Allen, S. E. Haupt and G. S. Young, 'Source Characterization with a Genetic Algorithm–Coupled Dispersion–Backward Model Incorporating SCIPUFF', *Journal of Applied Meteorology and Climatology*, 46, 3 (2007)

Bass, A., “Modeling long range transport and diffusion”, Second Joint conf. On Appl. Of Air Poll. Met., AMS & APCA, New Orleans, LA, pp 193-215 (1980)

K. C. Bowen and A. Stratton, “The Information Content of Data in Defined Choice-Situations” *The Journal of the Operational Research Society*, Vol. 32, No. 2 (1981)

Daescu, D. N., and Carmichael, G. R., “An Adjoint Sensitivity Method for the Adaptive Location of the Observations in Air Quality Modeling” *Journal of the Atmospheric Sciences*, 60, 2 (2003)

Donaldson, C. du P., “Atmospheric turbulence and the dispersal of atmospheric pollutants”, AMS Workshop on Micrometeorology, ed. D. A. Haugen, Science Press Boston, pp 313-390 (1973)

Errico, R.M., “What is an adjoint model”. *Bull. Am. Meteor. Soc.*, 78, 2577-2591 (1997)

R. Giering and T. Kaminski, 'Transformation of Algorithms in FORTRAN' (2006).

C.E. Kolb, S.C. Herndon, J.B. McManus, J.H. Shorter, M.S. Zahniser, D.D. Nelson, J.T. Jayne, M.R. Canagaratna, D.R. Worsnop, "Mobile Laboratory with Rapid Response Instruments for Real-Time Measurements of Urban and Regional Trace Gas and Particulate Distributions and Emission Source Characteristics", *Environ. Sci. Technol.*, 38, 5694-5703, (2004)

Lewellen, W. S., “Use of invariant modeling”, *Handbook of Turbulence*, ed. W. Frost and T. H. Moulden, Plenum Press, pp 237-280 (1977)

Lewellen, W. S., “Modeling the lowest 1km of the atmosphere”, AGARD, AGARD-AG-267 (1981)

K. J. Long, Haupt, S.E. and Young, G.S. “Assessing sensitivity of source term estimation”, *Atmospheric Environment*, Volume 44, Issue 12, (2010)

Lushi, E. and J. M. Stockie, “An inverse Gaussian plume approach for estimating atmospheric pollutant emissions from multiple point sources”, *Atmospheric Environment*, 44, 8 (2010)

Mavriplis, D. J., “Discrete Adjoint-Based Approach for Optimization Problems on Three-Dimensional Unstructured Meshes” *AIAA Journal* 45, 4 (2007)

N. Platt and D. Deriggi, “Comparative Investigation Of Source Term Estimation Algorithms Using Fusion Field Trial 2007 Data,” 8th Conference on Artificial Intelligence Applications to Environmental Sciences at AMS Annual Meeting, Atlanta, GA, Jan 17-21 (2010)

N. Platt, S. Warner and S. M. Nunes, 'Plan for Initial Comparative Investigation of Source Term Estimation Algorithms Using FUSION Field Trial 2007 (FFT 07)', Institute for Defense Analyses, IDA Document D-3488 (2008)

W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical recipes in C* (2nd ed.): the art of scientific computing, New York, NY, USA: Cambridge University Press, (1992)

K. S. Rao, 'Source estimation methods for atmospheric dispersion', *Atmospheric Environment*, 41 (2007)

M. Sharan, J.P. Issartel, S. Kumar Singh and P. Kumar, “An inversion technique for the retrieval of single-point emissions from atmospheric concentration measurements”, *Proc. R. Soc. A* 465, 2069–2088 (2009)

R. I. Sykes, W. S. Lewellen and S. F. Parker, 'A Gaussian Plume Model of Atmospheric Dispersion based on Second-Order Closure', *Journal of Climate and Applied Meteorology*, 25 (1985)

Sykes, R.I., W.S. Lewellen, and S.F. Parker, “A Gaussian Plume Model of Atmospheric Dispersion Based on Second-Order Closure”, *J. Clim. Appl. Met.*, 25, 322-331 (1986)

Sykes, R.I. and D.S. Henn, “Representation of Velocity Gradient Effects in a Gaussian Puff Model”, *Journal of Applied Meteorology*, volume 34, 2715-2723 (1995)

R. I. Sykes and R. S. Gabruk, 'A Second-Order Closure Model for the Effect of Averaging Time on Turbulent Plume Dispersion', *Journal of Applied Meteorology*, 36, 8 (1997)

Venkatram, A. and J. C. Wyngaard, “Lectures on air pollution modeling”, *American Meteorological Society*, Boston, MA (1988)

Wyngaard, J. C., "Structure of the Planetary Boundary Layer and Implications for its Modeling", J. Climate Appl. Meteor., 24, 1131-1142 (1985)

L. Zhang, E. M. Constantinescu, A. Sandu, Y. Tang, T. Chai, G. R. Carmichael, D. Byun and E. Olaguer, 'An adjoint sensitivity analysis and 4D-Var data assimilation study of Texas air quality', Atmospheric Environment, 42 (2008)